

SINGLE-INDEX REGRESSION MODELS

Jingwei Wu

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Biostatistics,
Indiana University

May 2015

Accepted by the Graduate Faculty, Indiana University, in partial
fulfillment of the requirements for the degree of Doctor of Philosophy.

Wanzhu Tu, Ph.D., Chair

Doctoral Committee

Hanxiang Peng, Ph.D.

Hai Liu, Ph.D.

March 31, 2015

J. Howard Pratt, M.D.

© 2015

Jingwei Wu

DEDICATION

To My Family.....

ACKNOWLEDGMENTS

I would like to express sincere gratitude to my thesis advisor, Dr. Wanzhu Tu, for his guidance and support of my PhD studies. He is not only a mentor in scholarship, but also a good friend who encouraged me to apply to the Biostatistics PhD program and worked with me until the completion of this degree. Throughout my dissertation research, he has invested a tremendous amount of time and effort to help me sort out the technical details of this work. His patience and support has helped me overcome many challenges to complete this dissertation. He set an example of professionalism and scholarly dedication.

I would especially like to thank Dr. Hanxiang Peng, another member on my thesis committee. Dr. Peng's high standard in research has motivated me to pursue the scientific excellence that this research requires.

I am grateful to Dr. Jaroslaw Harezlak, my preliminary exam committee member. He taught me longitudinal data analysis and semiparametric regression. The latter developed my curiosity and prompted me to consider the inclusion of nonparametrics in my work.

I am also indebted to other members on my preliminary exam and dissertation committee: Dr. Hai Liu, Dr. Howard Pratt, Dr. Terrel Zollinger and Dr. Kurt Kroenke. All have provided much insight into the potential utility of my work.

Further, I am thankful for the support from Department of Biostatistics at Indiana University. I am grateful to the faculty, staff members, and fellow students for their generous guidance and direction.

Last, but certainly not least, I would like to express genuine gratitude to my family: my parents, Dezu Wu and Guiyue Zhang, for their faithful love, support, and strength over the years; my beloved wife, Yi Sun, for her constant encouragement, unconditional love, timeless support, and loving prayers throughout this endeavor.

SINGLE-INDEX REGRESSION MODELS

Useful medical indices pose important roles in predicting medical outcomes. Medical indices, such as the well-known Body Mass Index (BMI), Charleson Comorbidity Index, etc., have been used extensively in research and clinical practice, for the quantification of risks in individual patients. However, the development of these indices is challenged; and primarily based on heuristic arguments. Statistically, most medical indices can be expressed as a function of a linear combination of individual variables and fitted by single-index model. Single-index model represents a way to retain latent nonlinear features of the data without the usual complications that come with increased dimensionality. In my dissertation, I propose a single-index model approach to analytically derive indices from observed data; the resulted index inherently correlates with specific health outcomes of interest. The first part of this dissertation discusses the derivation of an index function for the prediction of one outcome using longitudinal data. A cubic-spline estimation scheme for partially linear single-index mixed effect model is proposed to incorporate the within-subject correlations among outcome measures contributed by the same subject. A recursive algorithm based on the optimization of penalized least square estimation equation is derived and is shown to work well in both simulated data and derivation of a new body mass measure for the assessment of hypertension risk in children. The second part of this dissertation extends the single-index model to a multivariate setting. Specifically, a multivariate version of single-index model for longitudinal data is presented. An important feature of the proposed model is the accommodation of both correlations among multivariate outcomes and among the repeated measurements from the same subject via random effects that link the outcomes

in a unified modeling structure. A new body mass index measure that simultaneously predicts systolic and diastolic blood pressure in children is illustrated. The final part of this dissertation shows existence, root-n strong consistency and asymptotic normality of the estimators in multivariate single-index model under suitable conditions. These asymptotic results are assessed in finite sample simulation and permit joint inference for all parameters.

Wanzhu Tu, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xii
Chapter 1 Introduction	1
1.1 Motivations and Objectives	1
1.2 Dissertation Compendium	2
Chapter 2 A Single-Index Model for Longitudinal Data	5
2.1 Introduction	5
2.1.1 Background	5
2.1.2 Single-Index Models	7
2.2 Model Formulation and Parameter Estimation	9
2.2.1 A Partially Linear Single-Index Model for Longitudinal Data	9
2.2.2 A Model Fitting Procedure	10
2.2.3 Penalized Least Square Estimation	13
2.2.4 An Algorithm for Parameter Estimation	14
2.3 Simulation Study	15
2.4 Real Application: Development of a Pediatric Body Mass Index	17
2.4.1 Three Model Formulations for Height-Weight Relationship with BP	17
2.4.2 Description of Data Source	19
2.4.3 Results	19
2.5 Discussion	21
Chapter 3 A Multivariate Single-Index Model for Longitudinal Data	24
3.1 Introduction	24

3.2	A Multivariate Single-Index Model	25
3.2.1	Specification of A Multivariate Single-Index Model	25
3.2.2	Mixed Effect Model Representation	27
3.2.3	Estimation	30
3.2.4	Confidence Interval Estimate of the Mean Responses	31
3.3	Simulation	31
3.3.1	Data Generation	31
3.3.2	Simulation Results	33
3.4	Data Application	35
3.4.1	Different Index Functions of Height and Weight	36
3.4.2	Results	36
3.5	Discussion	38
3.6	Appendix	39
Chapter 4 Estimators in Multivariate Single-Index Models: Asymptotic Proper-		
	ties	42
4.1	Introduction	42
4.2	Statistical Model and Main Results	43
4.2.1	Multivariate Single-Index Model and Estimators	43
4.2.2	Main Results and Proofs	45
4.2.2.1	Asymptotics with $\{\lambda_{m;n} \rightarrow 0, \forall m = 1, \dots, M\}$	49
4.2.2.2	Asymptotics with $\{\lambda_{m;n}, \forall m = 1, \dots, M\}$ fixed	60
4.3	Simulations	62
4.3.1	Data Generation	62
4.3.2	Simulation Result	63
4.4	Discussion	65

Chapter 5 Conclusion	66
BIBLIOGRAPHY	69
CURRICULUM VITAE	

LIST OF TABLES

2.1	Summary of parameter estimates for simulation: true value of $(\alpha_1, \alpha_2) = \frac{1}{\sqrt{5}}(1, 2) = (0.4472, 0.8944)$	17
2.2	Summary of comparisons in three model fits	20
3.1	Summary of parameter estimates for simulation: true $(\alpha_1, \alpha_2) = \frac{1}{\sqrt{5}}(1, -2) = (0.4472, -0.8944), \beta_1 = 1.6, \beta_2 = 0.5, \beta_3 = -2.7, \rho_{12} = 0.25, \rho_{13} = 0.50, \rho_{23} = 0.75, \delta_2 = 0.8, \delta_3 = 0.6$	33
3.2	Summary of parameter estimates over 200 simulations for two positively and negatively correlated outcomes: true index functions: $\eta_1(\boldsymbol{\alpha}^T \mathbf{X}) = \exp(\boldsymbol{\alpha}^T \mathbf{X}), \eta_2(\boldsymbol{\alpha}^T \mathbf{X}) = \boldsymbol{\alpha}^T \mathbf{X}, N = 100, n_i = 10$, True $(\alpha_1, \alpha_2) = \frac{1}{\sqrt{5}}(1, -2) = (0.4472, -0.8944), \beta_1 = 1.6, \beta_2 = 0.5, \delta_2 = 0.8$	35
3.3	Summary of comparisons in two index model fits	37
4.1	Summary of parameter estimates over 200 simulation runs: true parameters $(\alpha_1, \alpha_2, \alpha_3) = \frac{1}{\sqrt{14}}(2, -1, 3) = (0.5345, -0.2673, 0.8018), \rho = 0.50, \delta = 0.9, \sigma = 2$	64

LIST OF FIGURES

2.1	Curve estimates and confidence bands for the simulated data. The dot-dashed curves are the corresponding 2.5% and 97.5% quantiles.	16
2.2	Curve estimates for the blood pressure cohort study	20
2.3	Curve estimates for the blood pressure cohort study, stratified by sex. (a) male children; (b) female children	21
3.1	Curve estimates and confidence bands for the simulated data with three correlated outcomes. The solid curves are the true mean functions; the dashed curves are the average cubic-spline fit over 200 simulations. The dot-dashed curves are the corresponding 2.5 % and 97.5 % quantiles.	34
3.2	Spline estimates and 95% confidence bands for the SBP and DBP measures, stratified by sex. Solid curves are for systolic blood pressure, dashed curves are for diastolic blood pressure.	38
4.1	Curve estimates and confidence bands for the simulated data with bivariate outcomes. The solid curves are the true mean functions; the dashed curves are the average cubic-spline fit over 200 simulations. The dot-dashed curves are the corresponding 2.5 % and 97.5 % quantiles.	65

Chapter 1

Introduction

1.1 Motivations and Objectives

Index measures are commonly used in medical research and clinical practice, mainly for the purpose of quantifying risks associated with certain disease outcomes. For example, body mass index (BMI) measures level of adiposity in individual patients; numerous studies have correlated BMI with adverse health outcomes. In practice, however, the construction of these indices almost always relies on heuristic arguments instead of analytical derivation. The general acceptance of the constructed index measure is contingent on the results of subsequent validation studies. In my dissertation, I propose an alternative approach of analytically deriving indices from observed data and directly linking the indices with the outcomes of interest. In doing so, the resulting indices possess analytically derived functional forms and can be correlated with disease outcomes.

From a statistical perspective, most of the indices that we use in medical research can be expressed as a function of a linear combination of individual variables $\eta(\boldsymbol{\alpha}^T \mathbf{X})$, often referred to as the single-index model. Under such a general formulation, for a given set of individual variables \mathbf{X} , one needs to determine the form of the index function η , as well as the coefficients for the individual variables, to determine the value of an index. Therefore, to derive an index measure associated with a particular health outcome Y , one only needs to find an appropriate function η and coefficient value $\boldsymbol{\alpha}$ such that $E(Y|\mathbf{X}) = \eta(\boldsymbol{\alpha}^T \mathbf{X})$.

Single-index models first emerge as a tool for dimension reduction (i.e. through an index function, a p -dimensional vector is reduced to a scalar while preserving the potentially nonlinear effect of the risk index.). By reducing a multidimensional independent variable

vector \mathbf{X} into a scalar *index* $\eta(\boldsymbol{\alpha}^T \mathbf{X})$, one hopes to retain much of the modeling flexibility through the nonlinear function $\eta(\cdot)$. In comparison with other nonlinear models, single-index models are easier to fit, and results are easier to interpret, especially if the index function $\eta(\cdot)$ is a monotone function.

One notable limitation of the existing single index models is the lack of accommodation of repeatedly measured data and multivariate data. Undoubtedly, inclusion of repeatedly measured observations or correlated multivariate data could substantially increase the computational challenge. It is perhaps because of these difficulties that single-index models have not been widely extended to longitudinal and multivariate settings. My dissertation focuses on these less developed areas. Specifically, the research objective of this dissertation is to propose a statistical approach based on three interrelated scientific topics: (1) development of medical indices that best quantify patient risks with longitudinal data; (2) derivation of indices that work for multiple outcomes; and (3) establishing a theoretical foundation for multivariate single-index models. These research questions, in addition to this dissertation's methodological novelty, have a broad range of practical impact in medical research and clinical practice.

1.2 Dissertation Compendium

To elucidate each objective, I have divided the dissertation into three parts. In the first part, I propose a partially linear single-index model for longitudinal data and develop a computationally efficient and stable estimation approach using P-spline. In other words, the smooth function $\eta(\cdot)$ is modeled by a penalized cubic spline which allows for a more flexible choice of knots and penalty as compared to other methods of smoothing splines. P-splines can be fit directly by the penalized nonlinear least squares method for which a

more snaggletoothed computational algorithm can be implemented in most of the existing statistical software.

In the second part of this dissertation, I extend the method to multivariate data settings and develop related procedures. Model specification is similar to the univariate case; however, more complicated and dynamic structures for the random subject effect vector and the random error vector are introduced and discussed. The random subject effect vector induces not only correlations among multiple outcomes, but also random processes among the repeated measurements from the same subject. The random error vector, at the same time, captures the underlying serial correlations within and between multiple stochastic processes. In contrast to the kernel smoothing approach, I propose to estimate the non-parametric index functions using P-spline basis functions which allow us to make inference on all model components within a mixed model representation. In this research, the cubic spline estimator of the index function and the coefficients of the regression models are obtained by minimizing a penalized weighted least-square estimation function. The random effect and random error are calculated using their conditional means given data.

In practice, all aforementioned analytical methods are used for the derivation of an adiposity index based on recorded height and weight data for prediction of blood pressure in children and adolescents. Various simulation results are presented.

In the last part of this dissertation, I examine the asymptotic properties of the parameter estimators in the multivariate model under a cross-sectional study design. I mainly focus on development of asymptotic properties of estimators based on the P-splines. Asymptotic properties, including the existence, strong consistency, and asymptotic normality of the estimators are derived under suitable conditions.

The dissertation is structured as follows: in Chapter 2, I present a partially linear single-index mixed effect model in a longitudinal data setting and explore an alternative

body mass measure for prediction of blood pressure in children. Chapter 3 extends a single-index model to situations of multiple outcomes and examines different forms of height and weight functions for predicting systolic and diastolic blood pressure in children. In Chapter 4, I present the asymptotic properties of estimators in the proposed multivariate model to aid parameter inference. The dissertation ends with a few concluding remarks in Chapter 5.

Chapter 2

A Single-Index Model for Longitudinal Data

In this chapter, I discuss the derivation of practically useful indices using longitudinal data. The derivation is based on a partially linear single-index model. To illustrate the process, I present a common index measure, one for measuring body mass index. I also present a recursive algorithm for the fitting of this mode.

2.1 Introduction

2.1.1 Background

Adiposity has long been recognized as a risk factor for hypertension (Masuo et al., 2000; Steinberger et al., 2009). Although the physiological mechanisms that modulate the effects of adiposity on blood pressure (BP) have not been fully elucidated, empirical data on obesity-hypertension are remarkably consistent across human populations (Davy and Hall, 2004). In most clinical studies, levels of adiposity are quantified by a simple index measure, the body-mass index (BMI), although the adiposity-BP association has also been examined for other adiposity measures including skinfold, waist circumference, neck circumference, and dual-energy X-ray (DEXA) (Cornier et al., 2011; Genton et al., 2002; Scherf et al., 1986; Sebo et al., 2008). Despite the heuristic origin of BMI, the current criteria for obesity diagnoses in pediatric populations are based on age and sex adjusted BMI percentile values (Mei et al., 2002; WHO, 2000).

From a measurement perspective, the greatest appeal of BMI is its simplicity. Based on two easily measurable quantities (height and weight), BMI can be calculated accurately in a wide variety of clinical and nonclinical settings. In research, BMI has been widely adopted

and its use has gained acceptance by the scientific community as a measure of human fatness (Gallagher et al., 1996; NHLBI, 1998). Since excess weight at a given height usually indicates disproportionate adipose tissue mass in the body, the current BMI formulation ($\text{weight}/\text{height}^2$) is a reasonable proxy for the amount of body fat. Indeed, high BMI values have been linked to various adverse health outcomes including increased risk for cardiovascular disease, chronic kidney disease, and metabolic syndromes such as type II diabetes (Cassano et al., 1990; Hadaegh et al., 2011; MacMahon et al., 1987; Stamler, 1991). Numerous studies reported strong and positive associations between BMI and BP (Dyer et al., 1999; Huang et al., 1998). Similar relationships were reported in children (Baker et al., 1999; Falkner et al., 2006). Interestingly, research also has shown that, in children, BP increases with height as well as with weight (Krauss et al., 1998; Shankar et al., 2005; Tu et al., 2009; Williams, 2008), thus raising a question about the wisdom of including height as a reciprocal component in the BMI formulation.

In this chapter, I try to determine the form of a height and weight function that is optimal for predicting BP in children. For this purpose, I develop a new statistical tool that belongs to a general class of models called single-index models. Using this new tool, I combine height and weight information into a one-dimensional index and link this index to BP. The main methodological contributions of this work include the extension of traditional single-index models to a longitudinal data setting and the development of related model fitting algorithms based on the penalized likelihood method. Although the method is developed in the context of body mass measurement and its relationship to blood pressure, the approach can be easily modified for applications in other clinical investigations.

2.1.2 Single-Index Models

The single-index model, in its original form, is an effort to retain latent nonlinear features of the data without complications of high dimensionality (Bellman, 1961). The generic formulation of a simple single-index model is as follows: $E(\mathbf{Y}|\mathbf{X}) = \eta(\boldsymbol{\alpha}^T \mathbf{X})$, where \mathbf{Y} is an outcome variable, \mathbf{X} are covariates of p dimension, \mathbb{R}^p , and $\boldsymbol{\alpha}$ is an unknown index parameter vector in \mathbb{R}^p . To ensure the identifiability of the model, one often assumes that the first element of $\boldsymbol{\alpha}$ is positive and $\|\boldsymbol{\alpha}\| = 1$. It is also assumed that $\eta(\cdot)$ is an estimable univariate index function. By reducing multidimensional independent variable vector \mathbf{X} into a single-dimensional “index” $\eta(\boldsymbol{\alpha}^T \mathbf{X})$, one hopes to retain much of the modeling flexibility through the nonlinear function $\eta(\cdot)$. In data analysis, the use of single-index models is almost always motivated by practical considerations. The ultimate motivations that underly the single-index parameter $\boldsymbol{\alpha}^T \mathbf{X}$ include the following: reducing the dimensionality of independent variable space in initial data exploration, easy interpretation of parameters under monotone link functions $\eta(\cdot)$, and computational convenience in estimating $\boldsymbol{\alpha}$ in a lower dimensional space for multivariable regression analysis (Carroll et al., 1997; Li, 1991).

Extending beyond a simple single-index model, the partially linear single-index model is a regression model consisting of a single-index component and a linear additive component (i.e. $E(\mathbf{Y}|\mathbf{X}, \mathbf{W}) = \eta(\boldsymbol{\alpha}^T \mathbf{X}) + \boldsymbol{\beta}^T \mathbf{W}$, where $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{W} \in \mathbb{R}^q$ are covariates of interest, and both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown model parameter vectors.). Adding parametric additive terms $\boldsymbol{\beta}^T \mathbf{W}$ to accommodate covariates not included in the nonlinear index function is important for practical purposes because, despite the recent advances in nonlinear regression methodology, testing of linear effects remains one of the mainstay inference practices in regression analysis. In my data application, the partially linear single-index modeling structure provides an ideal structure for developing body mass indice: incorporating the

height and weight influences in the single-index component while leaving other relevant factors to the additive terms.

In practice, fitting single-index models remains a challenge as there are no generally usable computational procedures available in common statistical software. However, a few general model fitting approaches have been discussed in literature, including semiparametric weighted least square methods (Härdle et al., 1993; Ichimura, 1993), the average derivatives estimation (Härdle and Stoker, 1989; Stoker, 1986), the minimum average variance estimation method (Xia and Härdle, 2006; Xia et al., 2002), the P-spline estimation method (Yu and Ruppert, 2002), and, more recently, a two-stage model fitting method (Wang et al., 2010; Wang and Yang, 2009). In this chapter, I extend the penalized spline estimation for the partially single-index model, as proposed by Yu *et al.* (2002), to a longitudinal setting. In particular, random effects are used to accommodate the appropriate variance-covariance structures in the responses; operationally, P-splines can be viewed as a mixed model so that both the smoothing parameter and index-component parameters are directly estimated by the restricted maximum likelihood method (Brumback et al., 1999).

The remainder of this chapter is organized as follows: section 2.2 presents a modeling structure that extends the partially linear single-index model to a longitudinal setting and discusses a generally usable computational algorithm. In Section 2.3, I assess the operating characteristics of the proposed method through a simulation study. Development of the new body mass index is presented in Section 2.4. I conclude this chapter with a few remarks in Section 2.5.

2.2 Model Formulation and Parameter Estimation

2.2.1 A Partially Linear Single-Index Model for Longitudinal Data

To begin, I present a generic partially linear single-index mixed effect model with a few added features. I consider a longitudinal data situation in which there are n subjects, and the i^{th} subject contributes n_i observations. Write $N = \sum_{i=1}^n n_i$. Let y_{ij} be the outcome from the i^{th} subject measured at time j , where $i = 1, \dots, n$ and $j = 1, \dots, n_i$. Let $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijd})^T$ be a d -dimension covariate vector measured from the i^{th} subject at time j . Furthermore, two fixed effects are included in the proposed model, where I define $\mathbf{w}_i^f = (w_{i1}^f, \dots, w_{iq_1}^f)^T$ as the q_1 -dimension vector of fixed baseline covariates (i.e. gender and other patient characteristics, treatment effect, etc.) for the i^{th} subject and $\mathbf{w}_{ij}^v = (w_{ij1}^v, \dots, w_{ijq_2}^v)^T$ as the time-varying (i.e. visit-specific) covariates with dimension q_2 . The time-varying covariates can contain the autoregressive components and can be used to capture the longitudinal contributions to the previously measured outcomes. Therefore, a single individual-specific random effect induces a correlation structure among repeated measures; the error vector $\boldsymbol{\epsilon}_i$ is assumed to be conditionally independent in the subject-specific mean of \mathbf{Y}_i . With this notation, the partially linear single-index mixed effect model can be written as follows:

$$\mathbf{Y}_i = \eta(\mathbf{X}_i \boldsymbol{\alpha}) + \mathbf{W}_i^f \boldsymbol{\beta}_f + \mathbf{W}_i^v \boldsymbol{\beta}_v + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n. \quad (2.1)$$

In this formulation, $\mathbf{Y}_i = (y_{i1}, \dots, y_{in_i})^T \in \mathbb{R}^{n_i}$ is the response vector for the i^{th} subject, $\mathbf{X}_i = (\mathbf{x}_{i1}^T, \dots, \mathbf{x}_{in_i}^T)^T \in \mathbb{R}^{n_i \times d}$ is a vector of index elements, and $\mathbf{W}_i^f = (\mathbf{w}_i^f, \dots, \mathbf{w}_i^f)^T \in \mathbb{R}^{n_i \times q_1}$ and $\mathbf{W}_i^v = (\mathbf{w}_{i1}^v, \dots, \mathbf{w}_{in_i}^v)^T \in \mathbb{R}^{n_i \times q_2}$ are subject-specific and time-varying covariate vectors, respectively. Function $\eta(\cdot)$ is an unknown univariate index function. Furthermore, $\mathbf{b}_i \in \mathbb{R}^r$ is a vector of random effects, $\mathbf{Z}_i \in \mathbb{R}^{n_i \times r}$, $r \leq q_1 + q_2$ is a known design matrix linking the vector of random effects \mathbf{b}_i to \mathbf{Y}_i , and $\boldsymbol{\epsilon}_i$ is a vector of errors, independent of

\mathbf{b}_i . Herein, $\boldsymbol{\epsilon}_i$ is assumed to follow a multivariate normal distribution $N(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n_i})$ although the variance-covariance matrix can be extended to a more general form if needed. Write the linear parameters as $\boldsymbol{\beta}_f \in \mathbb{R}^{q_1}, \boldsymbol{\beta}_v \in \mathbb{R}^{q_2}$ and the unknown single index parameters as $\boldsymbol{\alpha} \in \mathbb{R}^d$ (with constraint $\|\boldsymbol{\alpha}\|^2 = 1$ for identifiability).

Next, I propose to model the smooth function $\eta(\cdot)$ by a penalized spline (cubic-spline) which allows for a more flexible choice of knots and penalty as compared to other methods of smoothing splines (Ruppert and Carroll, 2000; Ruppert et al., 2002). P-splines can be fit directly by penalized nonlinear least squares for which a more snaggletoothed computational algorithm can be implemented in most of the existing statistical software.

2.2.2 A Model Fitting Procedure

To fit the partially linear single-index mixed effect model, I examine the random intercept model in the current application setting. Write the partially single-index mixed effect model with random subject effect as

$$\mathbf{Y}_i = \eta(\mathbf{X}_i \boldsymbol{\alpha}) + \mathbf{W}_i^f \boldsymbol{\beta}_f + \mathbf{W}_i^v \boldsymbol{\beta}_v + \mathbf{U}_i + \boldsymbol{\epsilon}_i, \quad (2.2)$$

where \mathbf{U}_i is the random subject effect. As proposed by Carroll et al. (1997), the unknown univariate function $\eta(\cdot)$ can be estimated by using a cubic-spline.

Let

$$\eta(u) = \gamma_0 + \gamma_1 u + \gamma_2 u^2 + \gamma_3 u^3 + \sum_{k=1}^K \gamma_{3+k} (u - \kappa_k)_+^3, \quad (2.3)$$

where $\{\kappa_k\}_{k=1}^K$ are the spline knots and $(u - \kappa_k)_+^3$ are truncated cubic functions.

Write the spline coefficient vector as $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_{3+K})^T$ and the spline basis as

$$\mathbf{B}(u) = \left(1, u, u^2, u^3, (u - \kappa_1)_+^3, \dots, (u - \kappa_K)_+^3\right). \quad (2.4)$$

Then, one has $\eta(u) = \mathbf{B}(u)\boldsymbol{\gamma}$. For notational convenience, write the fixed and time varying covariates as an $n_i \times q$ matrix of covariates, $\mathbf{W}_i = (\mathbf{W}_i^f, \mathbf{W}_i^v)$, where $q = q_1 + q_2$ column vectors correspond to q_1 time-invariant covariates and q_2 time-varying covariates.

With the notation, the proposed model can be written as

$$\mathbf{Y}_i = \mathbf{B}(\mathbf{u}_i)\boldsymbol{\gamma} + \mathbf{W}_i\boldsymbol{\beta} + \mathbf{U}_i + \boldsymbol{\epsilon}_i, \quad (2.5)$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ are model parameters to be estimated.

Let

$$\mathbf{X} = \begin{bmatrix} 1 & u_{11} & u_{11}^2 & u_{11}^3 & w_{111} & \dots & w_{11q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & u_{1n_1} & u_{1n_1}^2 & u_{1n_1}^3 & w_{1n_11} & \dots & w_{1n_1q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_{n1} & u_{n1}^2 & u_{n1}^3 & w_{n11} & \dots & w_{n1q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & u_{nn_n} & u_{nn_n}^2 & u_{nn_n}^3 & w_{nn_n1} & \dots & w_{nn_nq} \end{bmatrix}, \quad (2.6)$$

$$\mathbf{Z} = \begin{bmatrix} 1 & \dots & 0 & (u_{11} - \kappa_1)_+^3 & \dots & (u_{11} - \kappa_K)_+^3 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \dots & 0 & (u_{1n_1} - \kappa_1)_+^3 & \dots & (u_{1n_1} - \kappa_K)_+^3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & (u_{n1} - \kappa_1)_+^3 & \dots & (u_{n1} - \kappa_K)_+^3 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & (u_{nn_n} - \kappa_1)_+^3 & \dots & (u_{nn_n} - \kappa_K)_+^3 \end{bmatrix}, \quad (2.7)$$

and

$$\boldsymbol{\beta}^* = \begin{bmatrix} \gamma_0 \\ \vdots \\ \gamma_3 \\ \beta_1 \\ \vdots \\ \beta_q \end{bmatrix} \quad \mathbf{u}^* = \begin{bmatrix} U_1 \\ \vdots \\ U_n \\ \gamma_4 \\ \vdots \\ \gamma_{3+K} \end{bmatrix}. \quad (2.8)$$

The spline mixed effect model can then be represented in the usual mixed effects model,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}^* + \mathbf{Z}\mathbf{u}^* + \boldsymbol{\epsilon}, \quad \text{Cov} \begin{bmatrix} \mathbf{u}^* \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \sigma_U^2 \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\epsilon^2 \mathbf{I} \end{bmatrix}. \quad (2.9)$$

The best linear unbiased prediction (BLUP) of $(\boldsymbol{\beta}^*, \mathbf{u}^*)$ is given by

$$\begin{bmatrix} \tilde{\boldsymbol{\beta}}^* \\ \tilde{\mathbf{u}}^* \end{bmatrix} = (\mathbf{C}^T \mathbf{C} + \boldsymbol{\Lambda})^{-1} \mathbf{C}^T \mathbf{Y}, \quad (2.10)$$

where $\mathbf{C} \equiv [\mathbf{X}, \mathbf{Z}]$ and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_\epsilon^2 \text{Cov}(\mathbf{u}^*)^{-1} \end{bmatrix}. \quad (2.11)$$

The smoothing parameter $\lambda_u = \sigma_\epsilon / \sigma_u$ is usually selected via the restricted maximum likelihood (REML) estimation of the variance components σ_u^2 and σ_ϵ^2 (Ruppert et al., 2002).

2.2.3 Penalized Least Square Estimation

Define the fitted individual mean function

$$\hat{\mathbf{Y}}_i = \mathbf{B}(\hat{\mathbf{u}}_i) \hat{\boldsymbol{\gamma}} + \mathbf{W}_i \hat{\boldsymbol{\beta}} + \mathbf{U}_i. \quad (2.12)$$

The objective function for the penalized least square estimation is given by

$$R_{\lambda_u}(\boldsymbol{\theta}) = \left(\sum_{i=1}^n n_i \right)^{-1} \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 + \lambda_u^6 \hat{\boldsymbol{\gamma}}^T \mathbf{D} \hat{\boldsymbol{\gamma}}, \quad (2.13)$$

where \mathbf{D} is an appropriate positive semidefinite symmetric matrix and $\lambda_u \geq 0$ is a penalty parameter. In this research, \mathbf{D} is chosen to have a form in which its last K diagonal elements equal 1 and the rest equal 0. This means that only the sum of squares of the cubic-spline basis functions is penalized.

2.2.4 An Algorithm for Parameter Estimation

To implement, the following algorithm is proposed:

Step 1: Start with initial values $\hat{\boldsymbol{\alpha}}^{(0)}$. For example, ordinary least squares (OLS) estimates for the linear mixed effect model $\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{W}\boldsymbol{\beta} + \mathbf{U} + \boldsymbol{\epsilon}$ can be used. Normalize $\hat{\boldsymbol{\alpha}}^{(0)}$ such that $\|\hat{\boldsymbol{\alpha}}^{(0)}\| = 1$, and restrict the first element to be positive.

Step 2: Calculate the preliminary estimates of the index values $\{\mathbf{u}_i = (\mathbf{X}_i\hat{\boldsymbol{\alpha}}^{(0)}) : i = 1, \dots, n\}$. Then obtain the BLUP estimates of $\hat{\boldsymbol{\theta}}^{(0)} = (\hat{\boldsymbol{\alpha}}^{(0)T}, \hat{\boldsymbol{\beta}}^{(0)T}, \hat{\boldsymbol{\gamma}}^{(0)T})^T$.

Step 3: Obtain $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\gamma}}^T)^T$ iteratively by fixing the estimate of $\lambda_u^{(iter)}$ and by simultaneously minimizing the residual sum of square with respect to all components of $\boldsymbol{\theta}$ under the constraints $\|\boldsymbol{\alpha}\| = 1$ and $\alpha_1 > 0$. The iteration stops when $\|\boldsymbol{\theta}^{m+1} - \boldsymbol{\theta}^m\|$ converges to zero. The knots used for the basis functions depend on $\boldsymbol{\alpha}$ since they are sample quantiles on $\{\mathbf{X}_i\boldsymbol{\alpha} : i = 1, \dots, n\}$.

The fitted value \mathbf{Y} is then given by

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}^* + \mathbf{Z}\hat{\mathbf{u}}^*. \quad (2.14)$$

Thus the penalized least squares estimate of $\boldsymbol{\theta}$ minimizes $R_{\lambda_u}(\boldsymbol{\theta})$.

In order to add penalty term to the least square function, Yu et al. (2002) proposed to augment data as

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \hat{\mathbf{Y}} \\ \lambda_u^3 \mathbf{D}^{1/2} \hat{\boldsymbol{\gamma}} \end{bmatrix} \quad (2.15)$$

because

$$\left\| \begin{bmatrix} \mathbf{Y} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{Y}} \\ \lambda_u^3 \mathbf{D}^{1/2} \hat{\boldsymbol{\gamma}} \end{bmatrix} \right\|^2 = R_{\lambda_u}(\boldsymbol{\theta}). \quad (2.16)$$

Step 2 is achieved by using standard R function `lme()`, and step 3 is achieved by using function `optim()` to minimize R_{λ_u} . This strategy is effective because λ_u controls the smoothness of $\hat{\eta}$ but it has relatively little effect on $\hat{\alpha}$.

2.3 Simulation Study

I evaluate the performance of the proposed algorithm through a simulation study. In the simulation, the estimated parameters are compared with their true values, and the corresponding standard errors are reported. The variation of the estimation curve is captured by a confidence band for the mean. Each of the simulated data sets contains $n = 100$ subjects with $m_i = 10$ observations per subject over time. Data are generated from the following model:

$$y_{ij} = \exp(\alpha_1 x_{1ij} + \alpha_2 x_{2ij}) + \beta z_i + U_i + \epsilon_{ij},$$

where x_{1ij} and x_{2ij} are bivariate covariates with independent uniform $(0, 1)$ components, $z_i = 1$ with probability of 0.3, and $z_i = 0$ with probability of 0.7. The random effect U_i follows independent $N(0, \sigma_U^2)$, and the error term ϵ_{ij} follows independent $N(0, \sigma_e^2)$. Here, I take $\sigma_U^2 = 0.04$ and $\sigma_e^2 = 0.02$.

To examine the performance of the proposed model fitting procedure, different values are chosen for parameters α, β . Mean values of the parameter estimates are presented in a tabular form, and they are compared to the true parameter values under different sample sizes. In the simulation study, I use cubic splines with 20 knots. The knots are selected at equally spaced quantiles of the estimated index values u ; therefore, they will change iteratively when estimated index values are computed (Ruppert et al., 2002; Yu and Ruppert, 2002). The smoothing parameter λ_u is chosen by REML.

Figure 2.1 shows the average cubic-spline estimates over $N = 200$ simulated datasets and the corresponding 2.5% and 97.5% quantiles. As indicated by the figure, the average cubic-

spline fit using PLSIMEM algorithm correctly captures the shape of the true exponential function. In fact, the true curve is completely covered by the estimated confidence band. The 2.5% and 97.5% confidence limits are tightly around the average of the cubic-spline fit, showing relatively little variability in the estimation process.

Table 2.1 summarizes the results for the parameter estimates, including the sample mean, standard error, and smoothing parameter estimates based on the setting of true parameters and the size of simulated data sets. The simulation study indicates that the coefficient estimates for the index parameters are very close to their true values. For the categorical covariate, z , the coefficient estimates are satisfactory; the confidence bands for $z = 1$ is somewhat wider than that for $z = 0$, possibly reflecting the unbalanced sample sizes associated with Bernoulli probability $p = 0.3$. In summary, the proposed PLSIMEM fitting algorithm works effectively in fitting the data, as the parameter estimates are very close to the true values, and the standard error is reasonably small.

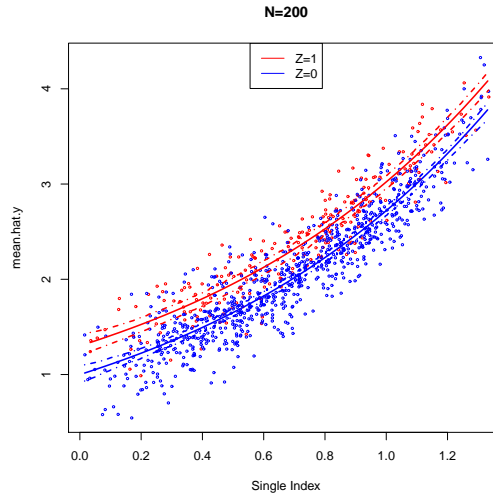


Figure 2.1: Curve estimates and confidence bands for the simulated data. The dot-dashed curves are the corresponding 2.5% and 97.5% quantiles.

Table 2.1: Summary of parameter estimates for simulation: true value of $(\alpha_1, \alpha_2) = \frac{1}{\sqrt{5}}(1, 2) = (0.4472, 0.8944)$

Parameter	N^a	$\beta=.1$		$\beta=.3$		$\beta=.5$	
		Mean	SE(10^{-3})	Mean	SE(10^{-3})	Mean	SE(10^{-3})
α_1	50	.4477	.997	.4477	.997	.4477	.998
	100	.4484	.676	.4483	.677	.4483	.677
	150	.4480	.582	.4480	.582	.4480	.582
	200	.4479	.508	.4479	.508	.4480	.531
α_2	50	.8941	.499	.8941	.499	.8941	.499
	100	.8938	.338	.8938	.338	.8938	.338
	150	.8940	.291	.8940	.291	.8940	.291
	200	.8940	.254	.8940	.254	.8940	.266
β	50	.1036	5.97	.3036	5.97	.5036	5.97
	100	.1047	4.53	.3047	4.53	.5047	4.53
	150	.1013	3.57	.3013	3.57	.5013	3.57
	200	.1015	3.05	.3015	3.05	.5015	3.29
λ	50	6.37		6.38		6.36	
	100	6.07		6.11		6.11	
	150	6.08		6.10		6.11	
	200	6.22		6.22		6.23	

^aN: Size of simulated datasets.

2.4 Real Application: Development of a Pediatric Body Mass Index

2.4.1 Three Model Formulations for Height-Weight Relationship with BP

In the context of single-index models, one could recover BMI from $f[\alpha_1 \log(\text{weight}) + \alpha_2 \log(\text{height})]$ by setting $\alpha_1 = 1$ and $\alpha_2 = -2$ and $f(\cdot) = \exp(\cdot)$. The question of interest is whether $f(\cdot) = \exp(\cdot)$ is the appropriate functional form and whether $\alpha_1 = 1$ and $\alpha_2 = -2$ are the most appropriate coefficient values for BP. A slight extension of the BMI formulation is to keep the functional form, but let coefficients α_1 and α_2 be determined by linear regression analysis. This second approach adds some flexibility as compared to BMI, but it forces a linear functional relationship between $\alpha_1 \log(\text{weight}) + \alpha_2 \log(\text{height})$ and BP. I present a method to optimize both functional form and coefficients of observed

weight and height via an index component for the purpose of deriving a simple and useful fatness measure that can be used to study the body mass relationship to BP.

Specifically, I write the three models as follows:

- Model 1: Linear mixed effect model for BMI (LMEM1):

$$SBP_{ij} = \beta_0 + \beta_1 \log BMI_{ij} + \beta^T W_{ij} + U_i + \epsilon_{ij} \quad (2.17)$$

- Model 2: Linear mixed effect model for height and weight (LMEM2):

$$SBP_{ij} = \beta_0 + \beta_1 \log WT_{ij} + \beta_2 \log HT_{ij} + \beta^T W_{ij} + U_i + \epsilon_{ij} \quad (2.18)$$

- Model 3: Partially Linear Single Index Mixed Effect Model (PLSIMEM):

$$SBP_{ij} = \eta(\alpha_1 \log WT_{ij} + \alpha_2 \log HT_{ij}) + \beta^T W_{ij} + U_i + \epsilon_{ij} \quad (2.19)$$

Directly utilizing the logarithm of BMI as a marker for body mass in relation to BP is represented in Model 1 which reflects the pre-defined weight/height² formulation. Model 2 provides the additive feature of observed weight and height measures by allowing different estimates for each logarithm measure of weight and height. Using the single-index approach, Model 3 provides a more flexible functional form of height-weight components than (2.18) for better prediction of BP.

To assess the performance of the proposed body mass index in Model 3 as compared to the two indices in the other two models, I examine the estimated values of the parameters as well as the magnitudes of integrated mean square errors.

2.4.2 Description of Data Source

The data used in this chapter come from a longitudinal observational study (Tu et al., 2011). Healthy children were recruited from schools in Indianapolis, Indiana. Enrolled children were followed longitudinally. During the course of follow-up, height, weight, and BP were measured twice a year. At the same time, each child's age and sex were documented. For this methodological exercise, I only use a subset of the study data, in which all subjects had at least 20 follow-up visits. To compare with the single-index model, I also consider two other aforementioned competing models. One used BMI as a predictor while the other used a linear combination of height and weight as a predictor.

2.4.3 Results

I use the data described in previous section to determine the best functional form of children's weight and height for prediction of longitudinally measured systolic blood pressure (SBP). The cohort used in the analysis included 103 children, 65 of whom were boys (63%). The mean age of the cohort at baseline was 13 years. All subjects had at least 20 follow-up visits. For the new body mass index, the logarithm of weight and height are considered as index components to be estimated, and age and sex are considered as fixed effects. A random subject effect is added to the model to incorporate the longitudinal measurements.

Figure 2.2 shows the PLSIMEM curve estimates at mean age of 13 years, stratified by sex. A non-linear curvature in $\hat{\eta}$ is noticeable. An interesting feature of this figure is the non-linearity curvature of $\hat{\eta}$ between 2.5 and 3.0, where the relatively high increment in SBP prediction is observed as compared to the other range of index values. The monotonic shape of the curve suggests that the single index values increase with SBP.

Table 2.2 summarizes model fitting results from the three indices (models) considered in the comparison. The coefficient estimates of logarithm of weight and height, in addition to

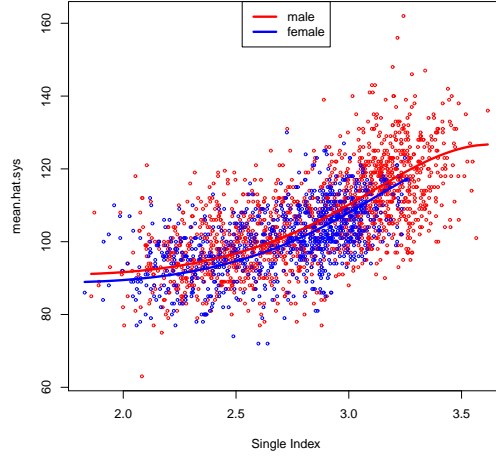


Figure 2.2: Curve estimates for the blood pressure cohort study

the fixed effects of age and sex, are given in the table. As mentioned previously, I normalize the estimated of logarithm of weight and height.

Table 2.2: Summary of comparisons in three model fits

Model	Parameter				ReMSE ^a
	logWT	logHT	Age	Male	
LMEM1	0.4472	-0.8944	0.4524	4.3622	89.8733
LMEM2	0.9195	-0.3930	-0.3289	2.9816	83.0125
PLSIMEM ^b	0.6389	0.7693	-0.4085	2.1636	53.2005

^a ReMSE: residual mean square error.

^b Penalized parameter estimate of λ_u is 0.4866 in the final model.

An important finding from the single-index model estimate is that the relative contributions of weight and height are rather different from the ratio of 1 to -2 as indicated in BMI; instead, they are approximately at a 1:1 ratio. This finding is consistent with previous studies, showing positive influences of height and weight on blood pressure (Shankar et al., 2005; Tu et al., 2009). This is an interesting finding and raises questions about the formulation of BMI, at least for the optimal prediction of blood pressure.

I further assess the goodness-of-fit of the models by examining the residual mean square error given in Table 2.2. Note that the residual mean square error of the new body mass

index derived from single-index model is 41% smaller than that of the traditional BMI. It also confirms that the PLSIMEM model fits the data much better than the two other competitors.

Finally, Figure 2.3 shows the PLSIMEM curve estimates with the 95% confidence band, stratified by sex. Figure 2.3a is for male and Figure 2.3b is for female, which clearly capture the non-linear feature of the *body mass-blood pressure* relationship. I conclude from Figure 2.3 that the single-index models fit the data adequately for both male and female children, and the new adiposity index has superior performance to BMI.

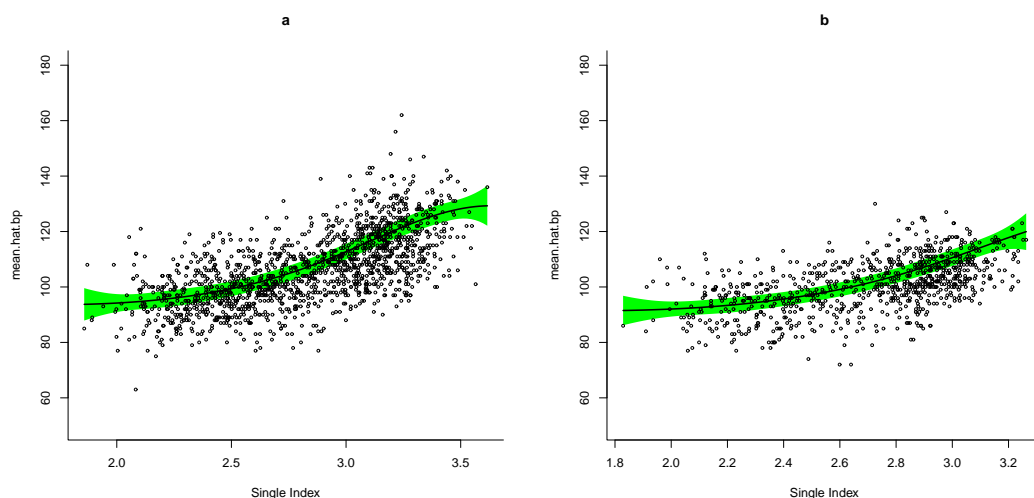


Figure 2.3: Curve estimates for the blood pressure cohort study, stratified by sex. (a) male children; (b) female children

2.5 Discussion

Motivated by a practical need for an optimal body mass index based on height and weight that can better predict blood pressure, I develop a general analytical tool for index function derivation. Using this new approach, I derive a pediatric body mass measure based on the observed weight and height for the prediction of BP. Unlike the empirically constructed BMI, the new index is derived from observed data using a modern statistical technique rather

than from a heuristic argument. I show that the resulting new index function of height and weight fitted longitudinally measured BP data much better than BMI. Interestingly, this finding indicates that both logarithm of children’s weight and height contributes to BP roughly the same. Similar techniques can potentially be used in the construction of other risk indices for other health outcomes.

From a methodological perspective, the use of single-index models is almost always motivated by practical considerations. However, despite theoretically attractive properties and the utility of the technique in risk index development, barriers remain when using this method. Fitting partially linear single index models, especially in longitudinal data settings, remains computationally challenging. This barrier is significant enough to discourage the use of this modeling approach in many suitable situations. Often times, the existing estimation methods are associated with heavy computational burden in high-dimensional spaces, possibly due to the use of more sophisticated nonlinear optimization techniques (i.e. simultaneous optimization of bandwidth choice and parameter estimation). More recently, studies have focused on the development of Bayesian computation methods (Anestis et al., 2004; Marley and Wand, 2010). But, the computational intensity and model instability associated with Bayesian method still remain formidable.

I have proposed a cubic-spline estimation scheme for partially linear single-index mixed effect models for data with repeated measurements. A recursive algorithm based on the optimization of penalized least square estimation equation is derived and shown to work well in both simulated and real data analysis. As previously recognized, an important advantage of using the penalized spline approach is the simultaneous estimation of both index parameter and index function with minimized computational complication. Single-index model analysis for the longitudinal or cluster data can be implemented using the proposed algorithm. The model fitting procedure is computationally efficient in practice.

Another innovation of this method is to use mixed effect model equivalence to choose the optimal smoothing parameter for the approximation of the index function. Many existing methods mostly rely on the generalized cross validation (GCV) technique for the identification of the smoothing parameter. Essentially, the GCV approach searches over a grid of possible penalty values and tends to be less efficient. In comparison, I compute the variance components of the random effect and estimation of error term to choose the optimal smoothing parameter used in the model, and this is computationally more efficient. With this approach, all parameters in the P-spline single-index model can be estimated simultaneously by the penalized nonlinear least squares.

In summary, the single-index modeling approach with the proposed estimation process provides a parsimonious structure with necessary flexibility to summarize high-dimensional data into a useful index with small computational burden. Such an extension represents a significant contribution to enhance both biomedical and epidemiological research. With the general and flexible modeling framework presented in this paper, I contend that the method has the potential to be used for the constructions of risk indices for other clinical outcomes. Along this line, I have proposed a practical method for combining multiple factors into a single usable index, and I have illustrated the use of the method through a real clinical application.

Chapter 3

A Multivariate Single-Index Model for Longitudinal Data

In this chapter, I extend the previously presented single-index model to a multivariate data setting. This extension is important because it ensures that the resulting index measure simultaneously works for multiple outcomes. The model is developed in the general framework of longitudinal data.

3.1 Introduction

An important factor contributing to the success of BMI is its ability to work with multiple outcomes. For all practical purposes, the utility of an index measure heavily depends on its ability to predict multiple outcomes.

In the previous chapters, I proposed a single-index model-based method to construct an index that correlates with a single outcome (Wu and Tu, 2013). I used truncated cubic lines as the basis for regression since their simple mathematical form is very useful when formulating complicated models (Eilers and Marx, 1996). I now extend the method to situations of multiple outcomes. This extension is practically important because no indices will be considered truly useful unless they work with multiple outcomes, as in the case of BMI.

The purpose of this chapter is to present a research tool that aids the development of index measures by directly linking the index functions to health outcomes through a multivariate single-index model. In presenting the method, I discuss the general form of the development model as well as related model fitting procedures. To illustrate, I examine different forms of height and weight functions, in comparison with the standard formulation

of BMI, for predicting systolic and diastolic blood pressure in children. Because of the limitations of the example data set, my discussion of BMI is solely for illustrative purposes and does not intend to propose a practical alternative to the widely accepted BMI.

3.2 A Multivariate Single-Index Model

3.2.1 Specification of A Multivariate Single-Index Model

Suppose we want to construct an index $\boldsymbol{\alpha}^T \mathbf{x}_{ij}$ based on the d -dimensional vector of independent variables $\mathbf{x}_{ij} \in \mathbb{R}^d$ and link the index function to M outcomes, $\tilde{\mathbf{Y}}_{ij} = (y_{1;ij}, \dots, y_{M;ij})^T$, for $i = 1, 2, \dots, N$ subjects, where each subject has $j = 1, \dots, n_i$ longitudinal observations.

The multivariate single-index model for each i^{th} subject as follows:

$$\left\{ \begin{array}{l} y_{1;ij} = \eta_1(\boldsymbol{\alpha}^T \mathbf{x}_{ij}) + \boldsymbol{\psi}_1^T \mathbf{w}_{ij} + b_{1;i} + \epsilon_{1;ij} \\ \vdots \\ y_{m;ij} = \eta_m(\boldsymbol{\alpha}^T \mathbf{x}_{ij}) + \boldsymbol{\psi}_m^T \mathbf{w}_{ij} + b_{m;i} + \epsilon_{m;ij} \\ \vdots \\ y_{M;ij} = \eta_M(\boldsymbol{\alpha}^T \mathbf{x}_{ij}) + \boldsymbol{\psi}_M^T \mathbf{w}_{ij} + b_{M;i} + \epsilon_{M;ij} \end{array} \right. \quad \forall i = 1, \dots, N, j = 1, \dots, n_i, \quad (3.1)$$

where $\boldsymbol{\alpha}$ is a $d \times 1$ vector of index coefficients for independent variables $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijd})^T$ which is common for all outcomes. In this model, I also include additional fixed effects $\mathbf{w}_{ij} = (w_{ij1}, \dots, w_{ijq})^T$ and the corresponding coefficient vectors $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_M \in \mathbb{R}^q$. To retain maximal flexibility, the index functions η_1, \dots, η_M are assumed to be outcome-specific, although in specific applications one may want to restrict the index functions to a common form across the outcomes. It is further assumed that the index functions are twice-differentiable smoothing functions. Finally, I include in the model random effects

$\tilde{\mathbf{B}}_i = (b_{1;i}, \dots, b_{M;i})^T$ and random errors $\tilde{\mathbf{\Xi}}_i = (\epsilon_{1;ij}, \dots, \epsilon_{M;ij})^T$, and I assume that $\tilde{\mathbf{B}}_i$ and $\tilde{\mathbf{\Xi}}_i$ are independent of each other.

Model (3.1) presents a system of simultaneous equations for M longitudinally measured outcomes. To link these equations in a unified structure, I use a random effect vector $\tilde{\mathbf{B}}_i$ and assume $\tilde{\mathbf{B}}_i$ follows $\text{MVN}(\mathbf{0}, \mathbf{\Sigma}_b)$. The diagonal elements of $\mathbf{\Sigma}_b$ are $\sigma_{b_m}^2$, and off-diagonal elements are $\rho_{st}\sigma_{b_s}\sigma_{b_t}$ for $m, s, t = 1, \dots, M$. Such a formulation gives us an intuitive interpretation: ρ_{st} is the correlation coefficient of the random subject effects for paired outcomes $y_{s;ij}$ and $y_{t;ij}$, whereas $\sigma_{b_s}^2$ and $\sigma_{b_t}^2$ are the corresponding variance components. Of note, this random subject effect vector induces not only a dependency structure among multivariate outcomes, but also correlations among the repeated measurements within the same subject. The random error vector, $\tilde{\mathbf{\Xi}}_i$, follows a zero-mean Gaussian distribution with variance-covariance matrix $\mathbf{\Sigma}_\epsilon$. Here $\mathbf{\Sigma}_\epsilon$ is a positive definite matrix whose components are determined by the underlying serial correlations within and between the M stochastic processes. For simplicity, I assume that the errors are independent although the structure could be relaxed to accommodate an autocorrelation structure. I write $\mathbf{\Sigma}_\epsilon$, which has a diagonal form $\mathbf{\Sigma}_\epsilon = \sigma_\epsilon^2 \text{diag}(1, \delta_2, \dots, \delta_M)$. Dispersion parameters $\sigma_{\epsilon_r}^2, r = 2, \dots, M$ are connected through scale parameters δ_r , so that $\sigma_{\epsilon_r}^2 = \delta_r \sigma_\epsilon^2$.

In the following sections, I use p-spline basis functions to estimate the nonparametric index functions, thus allowing presentation of the model in a mixed effect model format. I then obtain cubic spline estimates of index functions and regression coefficients by minimizing the weighted penalized least square functions. The random effects and random errors are calculated via best linear prediction and the restricted maximum likelihood (REML) method based on the observed data.

3.2.2 Mixed Effect Model Representation

Let $\mathbf{Y}_m = (y_{m;i1}, \dots, y_{m;in_i})_{1 \leq i \leq N}^T$, $\forall m = 1, \dots, M$ be the response vectors for m^{th} outcomes. Similarly, I denote $\mathbf{b}_m = (b_{m;1}, \dots, b_{m;N})$, $\mathbf{B}_m = (b_{m;1} \otimes \mathbf{1}_{n_1}^T, \dots, b_{m;N} \otimes \mathbf{1}_{n_N}^T)^T$ and $\mathbf{\Xi}_m = (\epsilon_{m;i1}, \dots, \epsilon_{m;in_i})_{1 \leq i \leq N}^T$, $\forall m = 1, \dots, M$ be the vectors of subject-specific random effects and random errors. Here, \otimes represents tensor-product, and $\mathbf{1}_{n_i}$ is a vector that contains n_i of element 1. Model (3.1) can therefore be written as

$$\mathbf{Y}_m = \eta_m(\mathbf{X}^* \boldsymbol{\alpha}) + \mathbf{W} \boldsymbol{\psi}_m + \mathbf{B}_m + \mathbf{\Xi}_m, \quad \forall m = 1, \dots, M,$$

where $\mathbf{X}^* = [\mathbf{x}_{ij}^T]_{1 \leq j \leq n_i; 1 \leq i \leq N}$ is the matrix of common index elements, and $\mathbf{W} = [\mathbf{w}_{ij}^T]_{1 \leq j \leq n_i; 1 \leq i \leq N}$ is the common subject-specific covariate matrix.

Let $\mathbf{v} = \mathbf{X}^* \boldsymbol{\alpha}$. Then the nonparametric index functions $\eta_m(\cdot)$ can be written as cubic splines:

$$\eta_m(v) = \gamma_{m;0} + \gamma_{m;1}v + \gamma_{m;2}v^2 + \gamma_{m;3}v^3 + \sum_{k=1}^K \gamma_{m;3+k}(v - \kappa_{m;k})_+^3,$$

where $\{\kappa_{m;k}\}_{k=1}^K$ and $(v - \kappa_{m;k})_+^3$ are the spline knots and truncated cubic function basis (K is the number of interpolate knots). The spline can be expressed as $\eta_m(v) = \mathbf{G}(v) \boldsymbol{\gamma}_m$, where $\boldsymbol{\gamma}_m = (\gamma_{m;0}, \dots, \gamma_{m;3+K})^T$ is the spline coefficient vector, and $\mathbf{G}(v) = \left(1, v, v^2, v^3, (v - \kappa_1)_+^3, \dots, (v - \kappa_K)_+^3\right)$ is the spline basis. It is well known that p-spline can be expressed in a mixed model representation by differentiating the un-penalized (fixed) and penalized (random) elements in $\mathbf{G}(v)$ formulation. Specifically, write $\mathbf{G}_F^i = [1, v_{ij}, v_{ij}^2, v_{ij}^3]_{1 \leq j \leq n_i}$ and $\mathbf{G}_R^i = [(v_{ij} - \kappa_1)_+^3, \dots, (v_{ij} - \kappa_K)_+^3]_{1 \leq j \leq n_i}$, so that $\mathbf{G}_F = [\mathbf{G}_F^i]_{1 \leq i \leq N}$ and $\mathbf{G}_R = [\mathbf{G}_R^i]_{1 \leq i \leq N}$ represent the “fixed” effects and “random” effects with corresponding parameter vectors $\boldsymbol{\gamma}_{Fm} = (\gamma_{m;0}, \dots, \gamma_{m;3})^T$ and $\boldsymbol{\gamma}_{Rm} = (\gamma_{m;4}, \dots, \gamma_{m;3+K})^T$. By combining the model fixed parameter vector $\boldsymbol{\psi}_m$ and subject-specific random effect vector \mathbf{b}_m , I have design matrices

$\mathbf{X} = [\mathbf{I}_M \otimes \mathbf{G}_F, \mathbf{I}_M \otimes \mathbf{W}]$ and $\mathbf{Z} = [\mathbf{I}_M \otimes \mathbf{1}_R, \mathbf{I}_M \otimes \mathbf{G}_R] = [\mathbf{Z}_B, \mathbf{Z}_R]$, where \mathbf{I}_M is an identity matrix of dimension M and $\mathbf{1}_R = \text{diag}[\mathbf{1}_{n_i}]_{1 \leq i \leq N}$.

Let $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_M^T)^T$, $\boldsymbol{\epsilon} = (\boldsymbol{\Xi}_1^T, \dots, \boldsymbol{\Xi}_M^T)^T$. The multivariate outcome model can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon},$$

where the fixed parameter vector $\boldsymbol{\beta} = (\boldsymbol{\gamma}_{Fm}^T, \boldsymbol{\psi}_m^T)_{1 \leq m \leq M}^T$ consists of model parameters representing both parametric and nonparametric components; the random effects vector $\mathbf{u} = (\mathbf{b}_m^T, \boldsymbol{\gamma}_{Rm}^T)_{1 \leq m \leq M}^T = (\mathbf{B}^T, \boldsymbol{\gamma}^T)^T$ contains parameters of random effects (includes penalized elements). The random effects \mathbf{u} follows a multivariate normal distribution $\text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_u)$, with $\boldsymbol{\Sigma}_u = \text{diag}(\boldsymbol{\Sigma}_b \otimes \mathbf{I}_N, \text{diag}(\boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_M)) = \text{diag}(\boldsymbol{\Sigma}_B, \boldsymbol{\Sigma}_\Gamma)$. Each $K \times K$ variance matrix, $\boldsymbol{\Gamma}_m$, controls the amount of smoothing in the estimation of index function of η_m , such that $\boldsymbol{\Gamma}_m \sim \text{MVN}(\mathbf{0}, \sigma_{\gamma_m} \otimes \mathbf{I}_N)$. The random error follows a multivariate normal distribution $\text{MVN}(\mathbf{0}, \mathbf{R})$, with $\mathbf{R} = \boldsymbol{\Sigma}_\epsilon \otimes \mathbf{I}_{N^*}$, where $N^* = \sum_{i=1}^N n_i$ be the number of total multivariate observations.

To calculate the index value \mathbf{v} , I need estimates of index parameters $\boldsymbol{\alpha}$. Here I impose constraints $\|\boldsymbol{\alpha}\| = 1$ and $\alpha_1 > 0$ to ensure parameter identifiability. For convenience, write the model parameters to be estimated as $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$. I estimate the variance component parameter vector $\boldsymbol{\tau} = (\sigma_{\gamma_m}, \sigma_{b_m}, \rho_{st}, \sigma_\epsilon, \delta_r)_{1 \leq m, s, t \leq M, 2 \leq r \leq M}^T$ by using the REML method. The random effects vector \mathbf{u} is obtained by using best predicted values.

Robinson (1991) described alternative ways to derive the best linear unbiased prediction (BLUP) solution to parameter $\boldsymbol{\beta}$. One simple, albeit ad hoc way, is to obtain “joint maximum likelihood estimates” of both fixed and random effects $\boldsymbol{\beta}$ and \mathbf{u} using *Henderson’s justification* (Henderson, 1975). In the linear mixed model framework, $\mathbf{Y}|\mathbf{u} \sim \text{MVN}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R})$, $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_u)$, and $[\mathbf{u}, \boldsymbol{\epsilon}]^T \sim \text{MVN}(\mathbf{0}, \text{diag}(\boldsymbol{\Sigma}_u, \mathbf{R}))$. Maximizing

the likelihood of the (\mathbf{Y}, \mathbf{u}) over unknown parameters $\boldsymbol{\beta}$ and \mathbf{u} leads to the criterion

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \mathbf{u}^T \boldsymbol{\Sigma}_u^{-1} \mathbf{u}. \quad (3.2)$$

For any given values of $\boldsymbol{\tau}$, the estimates of $\boldsymbol{\beta}$, \mathbf{B} , and $\boldsymbol{\gamma}$ are obtained by minimizing weighted penalized least square function,

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_B \mathbf{B} - \mathbf{Z}_R \boldsymbol{\gamma})^T \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_B \mathbf{B} - \mathbf{Z}_R \boldsymbol{\gamma}) + \mathbf{B}^T \boldsymbol{\Sigma}_B^{-1} \mathbf{B} + \boldsymbol{\gamma}^T \boldsymbol{\Sigma}_\Gamma^{-1} \boldsymbol{\gamma},$$

where $\boldsymbol{\Sigma}_B = (\boldsymbol{\Sigma}_b \otimes \mathbf{I}_N)$ is the variance-covariance matrix of the subject-specific random vector \mathbf{B} , and $\boldsymbol{\Sigma}_\Gamma = \text{diag}(\boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_M)$ is the penalized smoothing matrix of $\boldsymbol{\gamma}$.

This is equivalent to solving a weighted penalized least square problem:

$$\hat{\boldsymbol{\zeta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{B}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \underset{\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\gamma}}{\text{argmin}} \left\{ (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_B \mathbf{B} - \mathbf{Z}_R \boldsymbol{\gamma})^T \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_B \mathbf{B} - \mathbf{Z}_R \boldsymbol{\gamma}) \right. \quad (3.3)$$

$$\left. + \mathbf{B}^T \boldsymbol{\Sigma}_B^{-1} \mathbf{B} + \sum_{m=1}^M \lambda_m^{2p} \|\boldsymbol{\gamma}_{Rm}\|^2 \right\}.$$

When fitting a p th order spline, the smoothing parameters λ_m , control the amount of trade-off between goodness-of-fit of η_m and smoothness by imposing a penalty on the coefficients of $\boldsymbol{\gamma}_{Rm}$.

Mathematically, this is also equivalent to solving a generalized-weighted penalized least square problem:

$$\hat{\boldsymbol{\zeta}}^* = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \underset{\boldsymbol{\beta}, \boldsymbol{\gamma}}{\text{argmin}} \left\{ (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_R \boldsymbol{\gamma})^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_R \boldsymbol{\gamma}) + \sum_{m=1}^M \lambda_m^{2p} \|\boldsymbol{\gamma}_{Rm}\|^2 \right\}. \quad (3.4)$$

Here, $\mathbf{V} = \mathbf{Z}_B \boldsymbol{\Sigma}_B \mathbf{Z}_B^T + \mathbf{R}$ is a working covariance matrix depending on one or more parameters in $\boldsymbol{\tau}$ in the case of heteroscedastic and correlated errors. Compared to the linear mixed model representation, the smoothing parameters in the penalized weighted least square equation yield $\lambda_m^{2p} = 1/\sigma_{\gamma_m}^2$. See Appendix for a sketch of the derivation of this expression.

3.2.3 Estimation

I consider the computation of index components in a joint model setting with multivariate longitudinal data. The estimation process has the following three steps:

Step 1: Set initial values for index component parameters to $\hat{\boldsymbol{\alpha}}^{(0)}$. In the absence of information of the unknown parameters, use ordinary least square estimates obtained from linear mixed effect model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$ as initial values. To ensure model identifiability, normalize $\hat{\boldsymbol{\alpha}}^{(0)}$ such that $\|\hat{\boldsymbol{\alpha}}^{(0)}\| = 1$, and restrict the first element to a positive value.

Step 2: Given the values of $\hat{\boldsymbol{\alpha}}^{(0)}$, calculate index $\{\mathbf{v}_i^{(0)} = (\mathbf{X}_i^* \hat{\boldsymbol{\alpha}}^{(0)}) : i = 1, \dots, N\}$, and obtain BLUP estimates of $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\alpha}^{(0)T}, \boldsymbol{\beta}^{(0)T})^T$. Hence, the likelihood given by equation (3.2) is maximized over the parameter of interest. Equivalently, the penalized weighted least square value given by equation (3.3) is minimized and denoted as $\hat{L}^{(0)}$.

Step 3: Iteratively obtain $\boldsymbol{\theta}^{(k)}$ and $\hat{L}^{(k)}$ by updating the new index values $\mathbf{v}_i^{(k)}$ until $\hat{L}^{(k)} - \hat{L}^{(k-1)}$ converges to zero. This step involves the entire domain of $\boldsymbol{\alpha}$ in the optimization procedure. The knots used for the basis functions depend on $\boldsymbol{\alpha}$ since they are sample quantiles on $\{\mathbf{X}_i \boldsymbol{\alpha} : i = 1, \dots, N\}$.

Maximization of the likelihood function in Step 2 is implemented by using R function `lme()` with an appropriate design matrix and random component matrix. The standard `varFunc` classes included in the `nlme` library is used to model the heteroscedastic variance functions across multivariate outcome measurements (Pinheiro and Bates, 2000). Step 3

is achieved by using R function `optim()` to ensure the penalized weighted least squares function converges over the entire domain of α .

3.2.4 Confidence Interval Estimate of the Mean Responses

I construct confidence intervals for the mean responses. Suppose $\mathbf{C}_x = [\mathbf{X}_x, \mathbf{Z}_x]$ and $\hat{y}_x = \mathbf{X}_x \hat{\beta} + \mathbf{Z}_x \hat{\mathbf{u}} = \mathbf{C}_x \hat{\xi}$, where both $\hat{\beta}$ and $\hat{\mathbf{u}}$ are estimated BLUP of β and \mathbf{u} , and \hat{y}_x is the estimated BLUP of $y_x = \mathbf{X}_x \beta + \mathbf{Z}_x \mathbf{u}$. Variability of the predicted value can be written as $\text{var}[\hat{y}_x | \mathbf{u}] = [\mathbf{X}_x \ \mathbf{Z}_x] \text{Cov}([\hat{\beta}, \hat{\mathbf{u}}]^T | \mathbf{u}) [\mathbf{X}_x \ \mathbf{Z}_x]^T = \mathbf{C}_x \text{Cov}([\hat{\beta}, \hat{\mathbf{u}}]^T | \mathbf{u}) \mathbf{C}_x^T$, and the BLUP estimates of $(\tilde{\beta}, \tilde{\mathbf{u}})$ can be expressed as $(\mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} + \mathbf{\Lambda})^{-1} \mathbf{C}^T \mathbf{R}^{-1} \mathbf{y}$, where $\mathbf{C} \equiv [\mathbf{X}, \mathbf{Z}]$ and $\mathbf{\Lambda} = \text{diag}(\mathbf{0}, \Sigma_u^{-1})$.

Because $\text{Cov}(\mathbf{y} | \mathbf{u}) = \mathbf{R}$, I have $\text{Cov}([\hat{\beta}, \hat{\mathbf{u}}]^T | \mathbf{u}) = (\mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} + \mathbf{\Lambda})^{-1} \mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} (\mathbf{C}^T \mathbf{R}^{-1} \mathbf{C} + \mathbf{\Lambda})^{-1}$, which suggests that $\text{Cov}([\hat{\beta}, \hat{\mathbf{u}}]^T | \mathbf{u}) \cong (\mathbf{C}^T \hat{\mathbf{R}}^{-1} \mathbf{C} + \hat{\mathbf{\Lambda}})^{-1} \mathbf{C}^T \hat{\mathbf{R}}^{-1} \mathbf{C} (\mathbf{C}^T \hat{\mathbf{R}}^{-1} \mathbf{C} + \hat{\mathbf{\Lambda}})^{-1}$. As a result, $\widehat{\text{st.dev.}}[\hat{y}_x | \mathbf{u}] = \sqrt{\mathbf{C}_x \text{Cov}^{-1}([\hat{\beta}, \hat{\mathbf{u}}]^T | \mathbf{u}) \mathbf{C}_x^T}$. It then follows that an approximate $100(1 - \alpha)\%$ confidence interval is $\hat{y}_x \pm z_{1-\frac{\alpha}{2}} \cdot \widehat{\text{st.dev.}}[\hat{y}_x | \mathbf{u}]$.

3.3 Simulation

I conduct an extensive simulation study to investigate the finite sample performance of the proposed algorithm. The performance of the model fitting procedure is assessed by using the Monte Carlo method under different parameter settings. I examine the estimation accuracy and precision by assessing the bias and standard errors of the estimates. I also compare the estimated index function curves with the true index curves. Overall fitness of the model is characterized by mean square error (MSE).

3.3.1 Data Generation

I simulate a scenario involving three correlated outcome variables. In each simulation, data are generated from a trivariate normal distribution. The three true index functions are

$\eta_1(v) = e^v$, $\eta_2(v) = v$, and $\eta_3(v) = v^2$, which are chosen to represent both linear and nonlinear effects of the index on the outcomes. Specifically, the three outcome variables $(Y_{1;ij}, Y_{2;ij}, Y_{3;ij})$ are generated from the following models:

$$\begin{cases} y_{1;ij} = \exp(\alpha_1 x_{1ij} + \alpha_2 x_{2ij}) + \beta_1 z_i + b_{1i} + \epsilon_{1;ij}, \\ y_{2;ij} = (\alpha_1 x_{1ij} + \alpha_2 x_{2ij}) + \beta_2 z_i + b_{2i} + \epsilon_{2;ij}, \\ y_{3;ij} = (\alpha_1 x_{1ij} + \alpha_2 x_{2ij})^2 + \beta_3 z_i + b_{3i} + \epsilon_{3;ij}, \end{cases}$$

where index component covariates x_{1ij} and x_{2ij} are independently generated from Uniform $[0, 1]$.

Covariate z_i is obtained from a Bernoulli distribution with $\Pr(z_i = 1) = 0.3$ and $\Pr(z_i = 0) = 0.7$. Random subject effect $(b_{1i}, b_{2i}, b_{3i})^T \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma}_b)$, where $\mathbf{\Sigma}_b$ is the variance-covariance matrix with $\sigma_1^2 = 0.4$, $\sigma_2^2 = 0.2$, $\sigma_3^2 = 0.3$ and $\rho_{12} = 0.25$, $\rho_{13} = 0.50$, $\rho_{23} = 0.75$. These correlations are chosen to represent different strengths of the association (from relatively low to high) between the outcomes (Hinkle et al., 2003). Considering the heteroscedasticity of the three outcomes, random error vector $(\epsilon_{1;ij}, \epsilon_{2;ij}, \epsilon_{3;ij})^T$ is generated from $\text{MVN}(\mathbf{0}, \mathbf{\Sigma}_\epsilon)$. For simplicity, only independent errors are considered, with $\sigma_\epsilon^2 = 0.1$ and two scale parameters $\delta_2 = 0.8$ and $\delta_3 = 0.6$.

The simulation study considers four different sample sizes. Number of subjects varies between $N = 50$ and $N = 100$; each subject is assumed to have $n_i = 5$ or $n_i = 10$ observations. For each sample size setting, 200 datasets are generated. I use 20 knots to fit cubic spline models. For well behaved functions, such a number of knots is considered sufficient to ensure the desired flexibility (Crainiceanu et al., 2005). In each iterative step, the knots are computed and selected at equally spaced quantiles of the estimated index values v (Ruppert et al., 2002). The choice of the smoothing parameters λ_1 , λ_2 , and λ_3 involved in this procedure is based on mixed model representation and computed by the inverse of the estimated variability of truncated line functions.

3.3.2 Simulation Results

I compare the estimated values of parameters against the true values. The parameter estimation results, including the mean values of the parameter estimates (Mean), standard error (SE), bias and MSE, are summarized in Table 3.1. My simulation shows that the estimated coefficient values are close to the true values in all cases, and the standard errors of the estimates are generally small. In addition, both the correlated structures and heteroscedasticity of the three outcomes are correctly exhibited by these estimates. Not surprisingly, MSE of each parameter estimates also appear to decrease with increased number of subjects when the number of follow-up observations within each subject is fixed.

Table 3.1: Summary of parameter estimates for simulation: true $(\alpha_1, \alpha_2) = \frac{1}{\sqrt{5}}(1, -2) = (0.4472, -0.8944)$, $\beta_1 = 1.6, \beta_2 = 0.5, \beta_3 = -2.7, \rho_{12} = 0.25, \rho_{13} = 0.50, \rho_{23} = 0.75, \delta_2 = 0.8, \delta_3 = 0.6$

Parameter	n	$n_i = 5$				$n_i = 10$			
		Mean	SE	Bias	MSE(10^{-4})	Mean	SE	Bias	MSE(10^{-4})
α_1	50	0.4387	0.0030	-0.0085	0.8158	0.4414	0.0021	-0.0058	0.3837
	100	0.4462	0.0030	-0.0010	0.0721	0.4438	0.0015	-0.0034	0.1353
α_2	50	-0.8974	0.0014	-0.0030	0.1131	-0.8967	0.0010	-0.0023	0.0646
	100	-0.8941	0.0012	0.0003	0.0163	-0.8958	0.0007	-0.0014	0.0252
β_1	50	1.5998	0.0148	-0.0002	2.1926	1.6138	0.0135	0.0138	3.7421
	100	1.5912	0.0106	-0.0088	1.8852	1.5864	0.0097	-0.0136	2.8092
β_2	50	0.5061	0.0095	0.0061	1.2699	0.4936	0.0099	-0.0064	1.3924
	100	0.4959	0.0068	-0.0041	0.6325	0.5035	0.0071	0.0035	0.6224
β_3	50	-2.6948	0.0116	0.0052	1.6168	-2.7099	0.0119	-0.0099	2.3841
	100	-2.7053	0.0084	-0.0053	0.9869	-2.7049	0.0089	-0.0049	1.0286
ρ_{12}	50	0.2442	0.0107	-0.0058	1.4827	0.2566	0.0098	0.0066	1.3909
	100	0.2423	0.0071	-0.0077	1.0912	0.2533	0.0075	0.0033	0.6650
ρ_{13}	50	0.4974	0.0079	-0.0026	0.6924	0.5045	0.0077	0.0045	0.7885
	100	0.5045	0.0059	0.0045	0.5474	0.4963	0.0058	-0.0037	0.4709
ρ_{23}	50	0.7447	0.0055	-0.0053	0.5802	0.7421	0.0053	-0.0079	0.8977
	100	0.7496	0.0038	-0.0004	0.1430	0.7533	0.0032	0.0033	0.2155
δ_2	50	0.8047	0.0081	0.0047	0.8649	0.8075	0.0049	0.0075	0.8001
	100	0.8063	0.0059	0.0063	0.7406	0.8002	0.0038	0.0002	0.1449
δ_3	50	0.5947	0.0054	-0.0005	0.5717	0.6050	0.0040	0.0050	0.4152
	100	0.6013	0.0044	0.0013	0.2082	0.5980	0.0025	-0.0020	0.1023

Figure 3.1 depicts the average cubic-spline estimates for three correlated outcomes based on 200 simulated datasets and the corresponding 2.5% and 97.5% quantiles. As illustrated in the figure, the average cubic-spline fit obtained from the proposed procedure correctly captures both the nonlinear and linear features of the three true functions (exponential in Figure 1a, linear in Figure 1b, and squared in Figure 1c). This indicates that there is little bias between its fit and the true mean functions. The average integrated MSE over three fitted index functions is 7.24×10^{-2} ; the average integrated squared bias is 0.46×10^{-2} ; the average integrated variance is 6.78×10^{-2} . At the same time, both 2.5% and 97.5% quantiles are close to the true curves, showing very small variation in the estimates. Finally, notably wider confidence bands are observed for $z = 1$ compared to $z = 0$, reflecting the relative contributions from the categorical covariate, z , with Bernoulli probability $p = 0.3$.

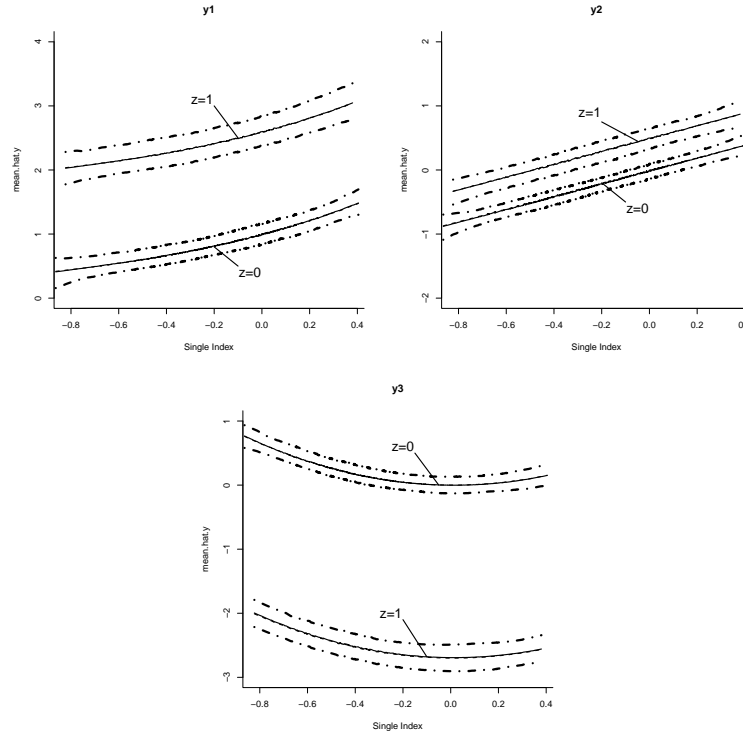


Figure 3.1: Curve estimates and confidence bands for the simulated data with three correlated outcomes. The solid curves are the true mean functions; the dashed curves are the average cubic-spline fit over 200 simulations. The dot-dashed curves are the corresponding 2.5 % and 97.5 % quantiles.

In summary, the simulation study shows that both the index components and curvature of index functions are accurately recovered. Other parameters associated with the multivariate linear models are also accurately estimated. The coverage probabilities of the confidence band are close to the nominal level, thus confirming that the proposed algorithm works well in tested data settings.

In the current simulated datasets, only positively correlated outcomes are considered. A separate simulation study evaluates the situation in which two negatively correlated outcomes are included. Again, the method performed well as expected, and the estimated correlations are close to the true values (See Table 3.2).

Table 3.2: Summary of parameter estimates over 200 simulations for two positively and negatively correlated outcomes: true index functions: $\eta_1(\boldsymbol{\alpha}^T \mathbf{X}) = \exp(\boldsymbol{\alpha}^T \mathbf{X})$, $\eta_2(\boldsymbol{\alpha}^T \mathbf{X}) = \boldsymbol{\alpha}^T \mathbf{X}$, $N = 100$, $n_i = 10$, True $(\alpha_1, \alpha_2) = \frac{1}{\sqrt{5}}(1, -2) = (0.4472, -0.8944)$, $\beta_1 = 1.6$, $\beta_2 = 0.5$, $\delta_2 = 0.8$.

Parameter	$\rho = 0.4$		$\rho = -0.4$	
	Mean	SE(10^{-2})	Mean	SE(10^{-2})
α_1	0.4465	0.0766	0.4453	0.1849
α_2	-0.8948	0.0382	-0.8949	0.0915
β_1	1.5828	0.6821	1.6137	0.8991
β_2	0.4977	0.4528	0.5019	0.7773
ρ	0.4023	0.9089	-0.3883	0.6309
δ_2	0.7936	0.5143	0.8050	0.3650

3.4 Data Application

To illustrate the proposed method, I use the longitudinal data described in the previous chapter to examine different formulations of height and weight functions for the prediction of systolic and diastolic blood pressure. I consider two different formulations: the standard BMI (weight/height²), which is known to correlate with blood pressure (Tu et al., 2009) and an index function derived from the proposed method.

3.4.1 Different Index Functions of Height and Weight

For the convenience of comparison, rewrite the BMI formulation as $f[\alpha_1 \log(\text{weight}) + \alpha_2 \log(\text{height})]$ with $\alpha_1 = 1$ and $\alpha_2 = -2$, where $f(\cdot) = \exp(\cdot)$. The alternative index takes the same form $f[\alpha_1 \log(\text{weight}) + \alpha_2 \log(\text{height})]$ without specifying the values of α_1 and α_2 . The functional form of $f(\cdot)$ is estimated from the observed data. Under this formulation, BMI is considered as a special case of the alternative index.

- Index 1: BMI

$$\begin{cases} \text{SBP}_{ij} = \beta_0^s + \beta_1^s \log \text{BMI}_{ij} + \beta_s^T \mathbf{W}_{ij} + \mathbf{U}_i^s + \epsilon_{ij}^s \\ \text{DBP}_{ij} = \beta_0^d + \beta_1^d \log \text{BMI}_{ij} + \beta_d^T \mathbf{W}_{ij} + \mathbf{U}_i^d + \epsilon_{ij}^d \end{cases}$$

- Index 2: An alternative function of height and weight

$$\begin{cases} \text{SBP}_{ij} = \eta_s(\alpha_1 \log \text{WT}_{ij} + \alpha_2 \log \text{HT}_{ij}) + \beta_s^T \mathbf{W}_{ij} + \mathbf{U}_i^s + \epsilon_{ij}^s \\ \text{DBP}_{ij} = \eta_d(\alpha_1 \log \text{WT}_{ij} + \alpha_2 \log \text{HT}_{ij}) + \beta_d^T \mathbf{W}_{ij} + \mathbf{U}_i^d + \epsilon_{ij}^d \end{cases}$$

By estimating the index functions and values of index coefficients, I show how an index can be developed so that it possesses an optimal functional form for the prediction of pre-specified outcomes. This analysis is based on a subset of the original study cohort, in which all subjects had at least 20 follow-up visits.

3.4.2 Results

The example dataset included 103 children (65 males). The mean age of the children at study entry was 13 years. Besides height and weight as index components, I also control for the effects of age and sex in the index model as fixed effects. A random subject effect is included to accommodate the within-subject correlation.

Model fitting results for the two different indices are presented in Table 3.3. While both indices are positively correlated with systolic and diastolic blood pressure, the weight and height contributions in the new index are approximately at 1:2 ratio, instead of the 1:-2 ratio specified by the standard BMI. The new index has a smaller value of the residual mean squared error (ReMSE=62.0108) than does BMI (ReMSE=65.8979). Using a modified R^2 for mixed effect models developed by Xu (Xu, 2003), I assess the proportions of variations explained by each index. The BMI model has an R^2 value of 0.40 while the new index has a R^2 value of 0.46, representing an 15% increase in the proportion of variation explained.

Table 3.3: Summary of comparisons in two index model fits

Model	Parameter						ReMSE ^c	R^{2d}
	logWT	logHT	Age _s ^a	Male _s ^a	Age _d ^b	Male _d ^b		
Index 1	0.4472	-0.8944	0.3620	4.2524	0.1561	1.0020	65.89785	0.40
Index 2 ^e	0.4863	0.8738	-0.3912	2.1035	-0.2348	-0.1064	62.01079	0.46

^a Parameter estimates for SBP model

^b Parameter estimates for DBP model

^c ReMSE: residual mean square error.

^d R^2 : proportions of explained variation.

^e Penalized parameter estimates of λ for SBP and DBP are 0.375 and 0.813, respectively.

The associations between the values of the new index and systolic and diastolic blood pressure in male and female subjects are graphically presented in Figure 3.2. The 95% pointwise confidence band, stratified by sex, are quite narrow, suggesting less variability and excellent precision of the new estimates.

In summary, I have shown that the new index development tool can be used to analytically derive useful indices, with greater R^2 values and greater proportions of outcome variability being explained. This said, the example is purely for demonstration purposes. The limited dataset, by itself, does not establish the superiority of the new index over the standard BMI.

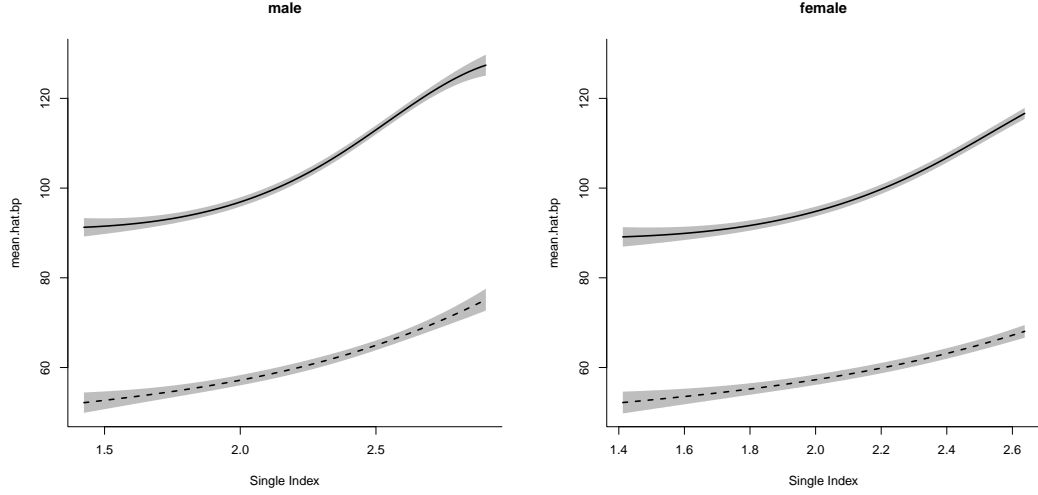


Figure 3.2: Spline estimates and 95% confidence bands for the SBP and DBP measures, stratified by sex. Solid curves are for systolic blood pressure, dashed curves are for diastolic blood pressure.

3.5 Discussion

Derivation of useful medical indices that correlate with multiple health outcomes is an issue of great practical importance. In this chapter, I provide a new tool to achieve this goal. By extending the partially linear single-index model to a multivariate setting, I develop a multivariate single-index model that allows investigators to analytically derive clinical indices that work for multiple clinical outcomes.

In addition to model construction, I develop the basic construction of the index development model, as well as related model fitting procedures. Simulation study shows the new method has excellent performance in estimation accuracy and computational efficiency. The model formulation is quite general and can accommodate longitudinal measures of multiple outcomes. Besides the index function, the model also includes other fixed and random effects. The index function is modeled by cubic splines and is estimated by using the penalized least square method. As I show in the simulation studies, both index components and curvature of the index functions are recovered accurately. The relatively narrow confi-

dence bands associated with fitted curves further attest to the model's estimation efficiency. Finally, as a index development tool, the method can be implemented in most computing platforms with existing software, thus has the potential to be used by practitioners in wide variety of applications.

3.6 Appendix

Consider a general multivariate p-spline model with K knots:

$$y_{m;i} = f_m(x_i) + \epsilon_{m;i}, \quad i = 1, \dots, N; \forall m = 1, \dots, M,$$

where the i^{th} response of the m^{th} outcome variable $y_{m;i}$ is generated from function f_m evaluated at covariate x_i plus a random error $\epsilon_{m;i}$. Assume $\epsilon_{m;i} \sim i.i.d.N(0, \sigma_\epsilon^2 \delta_m)$ for each fixed m . I estimate each smooth function f_m by a penalized spline with degree of p , that is:

$$f_m(x_i) = \beta_0^m + \beta_1^m x_i + \dots + \beta_p^m x_i^p + \sum_{k=1}^K \beta_{pk}^m (x_i - \kappa_k)_+^p. \quad (3.5)$$

The basis using truncated p -power functions are initiated for each sets of the outcome variable. I consider fitting the p th-degree spline by penalized least squares, whereas only non-polynomial coefficients are penalized. Denote $\mathbf{y}_m = (y_{m;1}, \dots, y_{m;N})^T$, $\mathbf{x}_i = (1, x_i, \dots, x_i^p)$, $\mathbf{z}_i = ((x_i - \kappa_1)_+^p, \dots, (x_i - \kappa_K)_+^p)$, $\mathbf{X}_m = (\mathbf{x}_1^T, \dots, \mathbf{x}_N^T)^T$, $\mathbf{Z}_m = (\mathbf{z}_1^T, \dots, \mathbf{z}_N^T)^T$, $\boldsymbol{\beta}_m = (\beta_0^m, \dots, \beta_p^m)^T$, $\boldsymbol{\gamma}_m = (\beta_{p1}^m, \dots, \beta_{pK}^m)^T$, $\boldsymbol{\zeta}_m = (\boldsymbol{\beta}_m^T, \boldsymbol{\gamma}_m^T)^T$. Further define $\mathbf{Y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_M^T)^T$, $\mathbf{C} = (\mathbf{I}_M \otimes \mathbf{X}_m, \mathbf{I}_M \otimes \mathbf{Z}_m) \equiv (\mathbf{X}, \mathbf{Z})$, $\boldsymbol{\zeta} = [(\boldsymbol{\beta}_m^T)_{1 \leq m \leq M}, (\boldsymbol{\gamma}_m^T)_{1 \leq m \leq M}]^T$, so that the weighted penalized least squares fit can be written as

$$\text{minimize } (\mathbf{Y} - \mathbf{C}\boldsymbol{\zeta})^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{C}\boldsymbol{\zeta}) \quad \text{subject to } \boldsymbol{\zeta}_m^T \mathbf{D} \boldsymbol{\zeta}_m < d_m.$$

Herein, \mathbf{V} is the heterostatic random errors matrix, with $\mathbf{V} = \mathbf{\Sigma}_\epsilon \otimes \mathbf{I}_N$, and $\mathbf{\Sigma}_\epsilon$ is the variance-covariance matrix, such that the diagonal elements are $\sigma_{\epsilon_1}^2, \dots, \sigma_{\epsilon_M}^2$, and off-diagonal elements are the pairwise covariances. Matrix \mathbf{D} has a specific structure,

$$\mathbf{D} = \begin{bmatrix} \mathbf{0}_{(p+1) \times (p+1)} & \mathbf{0}_{(p+1) \times K} \\ \mathbf{0}_{K \times (p+1)} & \mathbf{I}_{K \times K} \end{bmatrix},$$

which corresponds to the constraints on γ_m , so that $\|\gamma_m\|^2 < d_m$.

Using Lagrange multiplier method, the above minimization is equivalent to finding the minimized ζ of

$$(\mathbf{Y} - \mathbf{C}\zeta)^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{C}\zeta) + \lambda_1^{2p} \zeta_1^T \mathbf{D} \zeta_1 + \dots + \lambda_M^{2p} \zeta_M^T \mathbf{D} \zeta_M$$

for some $\lambda_m > 0$. In other words, each term $\lambda_m^{2p} \zeta_m^T \mathbf{D} \zeta_m$ penalizes the fit of function f_m associated with m^{th} outcome \mathbf{y}_m . This is equivalent to solving the following minimization problem with respect to ζ ,

$$(\mathbf{Y} - \mathbf{C}\zeta)^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{C}\zeta) + \zeta^T \mathbf{\Lambda} \zeta,$$

where $\mathbf{\Lambda}$ has a block-diagonal form, $\mathbf{\Lambda} = \text{diag}(0 \cdot \mathbf{1}_{M(p+1)}, \lambda_1^{2p} \cdot \mathbf{1}_K, \dots, \lambda_M^{2p} \cdot \mathbf{1}_K)$.

This has the solution,

$$\hat{\zeta} = (\mathbf{C}^T \mathbf{V}^{-1} \mathbf{C} + \mathbf{\Lambda})^{-1} \mathbf{C}^T \mathbf{V}^{-1} \mathbf{Y}. \quad (3.6)$$

On the other hand, if I assume $\beta_{pk}^m \sim N(0, \sigma_{\gamma_m}^2)$, the aforementioned multivariate p-spline model (3.5) corresponds to a framework of mixed effect model. Considering the best

linear unbiased prediction in the mixed model criterion, I have

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = (\mathbf{C}^T \mathbf{V}^{-1} \mathbf{C} + \mathbf{Q})^{-1} \mathbf{C}^T \mathbf{V}^{-1} \mathbf{Y}, \quad (3.7)$$

where $\mathbf{C} \equiv (\mathbf{X}, \mathbf{Z})$ and

$$\mathbf{Q} \equiv \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\gamma}^{-1} \end{bmatrix}.$$

Indeed, $\text{Cov}(\boldsymbol{\gamma}) = \left[\text{blockdiagonal}_{1 \leq m \leq M} \boldsymbol{\Sigma}_{\gamma_m} \right]$, where $\boldsymbol{\Sigma}_{\gamma_m} = \text{diag}(\sigma_{\gamma_m}^2 \cdot \mathbf{1}_K)$ and $\boldsymbol{\gamma}_m \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma_m})$. Therefore, if I treat $\lambda_m^{2p} = 1/\sigma_{\gamma_m}^2$, the solution of equation (3.6) and (3.7) are equivalent.

Extending to a random effect model, where the subject-specific random vector \mathbf{B} has variance-covariance matrix $\boldsymbol{\Sigma}_B = (\boldsymbol{\Sigma}_b \otimes \mathbf{I}_N)$, will not change the equivalence relationship between λ_m^{2p} and $\sigma_{\gamma_m}^2$. This is quite simplistic and straightforward because such a model assumes

$$\text{Cov}(\boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{B}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\gamma}^{-1} \end{bmatrix}.$$

The block-diagonal structure ensures the invariance relationship of $\lambda_m^{2p} = 1/\sigma_{\gamma_m}^2$ in the random effect model.

Chapter 4

Estimators in Multivariate Single-Index Models: Asymptotic Properties

This chapter examines the theoretical properties of estimators in multivariate single-index models. The existence, consistency, and asymptotic normality of all parameter estimators are derived under suitable regularity conditions. This development provides a theoretical foundation for large sample inferences concerning parameters of interest.

4.1 Introduction

The purpose of the current chapter is to examine the theoretical properties of the single-index model parameters. For univariate single-index models, a number of authors have studied the properties of single-index models when kernel smoothing and the empirical rule for bandwidth selection are used (Xia and Hardle, 2006; Xia et al., 2012, 2002; Zhang, 2007; Zhu and Xue, 2006). Though theoretically appealing, obtaining a kernel estimate for the marginal density of an index parameter can be challenging (Yu and Ruppert, 2002). Alternatively, there is a large literature on asymptotic properties of estimators in univariate single-index model, where the index function is estimated by the penalized spline (Bai et al., 2009; Tian et al., 2010; Yu and Ruppert, 2002, 2004). Yu and Ruppert (2002) showed that estimators obtained from the penalized spline have the desired \sqrt{n} -consistency property under the assumption of a compact parameter space Θ of parameter θ from cross-sectional data. Asymptotic properties, including the existence, strong consistency, and asymptotic normality of the estimators for longitudinal data are proved under suitable conditions (Bai et al., 2009; Tian et al., 2010). However, no studies, to the best of my knowledge, have explored such asymptotic properties in multivariate single-index models. Herein, I focus

on development of asymptotic properties of estimators based on the P-spline approach as discussed previously. Specifically, I study the properties of estimators in cross-sectional multivariate single-index models. It is important to note that similar theoretical conclusions can be extended to longitudinal settings. In fact, constructing a cross-sectional multivariate single-index model is equivalent to fitting a longitudinal single-index model, among which the multiple measure quantities are correlated. Therefore, the asymptotics derived from a longitudinal single-index model can be directly extended to a multivariate single index model under suitable conditions (Tian et al., 2010).

4.2 Statistical Model and Main Results

4.2.1 Multivariate Single-Index Model and Estimators

Let $y_{m;i}$ denote the m^{th} outcome ($m = 1, \dots, M$) for the i^{th} individual ($i = 1, \dots, n$).

Without loss of generality, consider the multivariate single-index model,

$$y_{m;i} = f_m(\mathbf{x}_{m;i}^T \boldsymbol{\alpha}) + \epsilon_{m;i}, \quad i = 1, \dots, n; \forall m = 1, \dots, M, \quad (4.1)$$

where $\boldsymbol{\alpha} \in \mathbb{R}^d$ is vector of index parameters associated with index covariates $\mathbf{x}_{m;i} \in \mathbb{R}^d$, and $\epsilon_{m;i}$ is a random error with mean 0 and variance σ_m^2 . The constraints $\|\boldsymbol{\alpha}\| = 1$ and $\alpha_1 > 0$ are imposed to ensure the identifiability. The unknown function $f_m(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is a twice-differentiable smoothing function for m^{th} outcome. Denote $\boldsymbol{\epsilon}_i = (\epsilon_{1;i}, \dots, \epsilon_{M;i})^T$. I assume $\boldsymbol{\epsilon}_i$ is dependent of \mathbf{x}_i and $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}_i$. I omit the additive term of $\boldsymbol{\beta}_m^T \mathbf{z}_{m;i}$ here because the same proof is valid for a partially linear multivariate single-index model if I expand the parameter space and corresponding design matrices.

To further simplify, define indicator functions $\mathcal{I}_l(m), \forall l = 1, \dots, M$ for each m^{th} outcome by

$$\mathcal{I}_l(m) = \begin{cases} 1 & \text{if } l = m \\ 0 & \text{if } l \neq m \end{cases},$$

so the equation (4.1) becomes

$$y_{m;i} = \mathcal{I}_1(m)f_1(\mathbf{x}_{m;i}^T\boldsymbol{\alpha}) + \mathcal{I}_2(m)f_2(\mathbf{x}_{m;i}^T\boldsymbol{\alpha}) + \dots + \mathcal{I}_M(m)f_M(\mathbf{x}_{m;i}^T\boldsymbol{\alpha}) + \epsilon_{m;i}. \quad (4.2)$$

I propose to estimate each of the unknown smoothing function $f_m(\cdot)$ by different penalized spline (p order spline). Specifically, if I define

$$f_m(v) = \sum_{q=1}^p \beta_{m;q} v^q + \sum_{k=1}^K \beta_{m;p+k} (v - \kappa_{m;k})_+^p, \quad \forall m = 1, \dots, M,$$

where $\{\kappa_{m;k}\}_{k=1}^K$ and $(v - \kappa_{m;k})_+^p$ are the p-spline knots and truncated function basis (K is the number of interpolate knots). Then the spline for m^{th} outcome measures can be expressed as $f_m(v) = \boldsymbol{\beta}_m^T \mathbf{G}(v)$. Here, $\boldsymbol{\beta}_m = (\beta_{m;0}, \dots, \beta_{m;p+K})^T$ represents the spline coefficient vector and spline basis is expressed as $\mathbf{G}(v) = \left(1, v, \dots, v^p, (v - \kappa_1)_+^p, \dots, (v - \kappa_K)_+^p\right)^T$. Define $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_M^T)^T$, the mean function for each m^{th} outcome measures, becomes

$$m_{m;i}(\boldsymbol{\theta}) \triangleq m_m(\mathbf{x}_{m;i}; \boldsymbol{\theta}) = \boldsymbol{\beta}^T \mathbf{B}(\mathbf{x}_{m;i}^T \boldsymbol{\alpha}), \quad (4.3)$$

where $\mathbf{B}(v) = \left(\mathbf{G}(v)^T \cdot \mathcal{I}_1(m), \dots, \mathbf{G}(v)^T \cdot \mathcal{I}_M(m)\right)^T$.

Therefore, the weighted penalized spline least square estimator of $\boldsymbol{\theta}$ minimizes

$$\begin{aligned} Q_{n,\lambda_1,\dots,\lambda_M}(\boldsymbol{\theta}) &= Q_n(\boldsymbol{\theta}) + \sum_{m=1}^M \lambda_m \boldsymbol{\beta}_m^T \mathbf{D}_m \boldsymbol{\beta}_m \\ &= \frac{1}{n} \sum_{i=1}^n \left(\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\theta}) \right)^T \mathbf{W}_i^{-1} \left(\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\theta}) \right) + \sum_{m=1}^M \lambda_m \boldsymbol{\beta}_m^T \mathbf{D}_m \boldsymbol{\beta}_m. \end{aligned}$$

Here, $\mathbf{y}_i = (y_{1;i}, \dots, y_{M;i})^T$ is a vector representation of M outcome measures for i^{th} subject, $\mathbf{m}_i = (m_{1;i}^T, \dots, m_{M;i}^T)^T$ is vector-based mean function for i^{th} subject, and \mathbf{W}_i is symmetric and positive-definite working covariance matrix. In addition, each \mathbf{D}_m is an appropriate positive semidefinite symmetric matrix, and $\{\lambda_m \geq 0, m = 1, \dots, M\}$ are penalty parameters. In this p-spline problem, for example, each \mathbf{D}_m is chosen to be diagonal with its last K diagonal elements equal to 1 and the rest equal to 0. In other words, it penalizes the sum of squares of the parameters in the p^{th} degree (Ruppert et al., 2002). All parameters can be estimated explicitly by minimizing $Q_{n,\lambda_1,\dots,\lambda_M}(\boldsymbol{\theta})$ via nonlinear least squares, and the selection of λ_m can be done by either generalized cross validation (GCV) or by means of mixed model representation (Ruppert et al., 2002; Wu and Tu, 2012; Yu and Ruppert, 2002).

4.2.2 Main Results and Proofs

Before obtaining the asymptotic properties, I begin by specifying the identifiability constraints $\|\boldsymbol{\alpha}\| = 1$ and $\alpha_1 > 0$ on the d -dimensional single-index parameter $\boldsymbol{\alpha}$ by reparameterization. Following Yu and Ruppert (2002), let $\boldsymbol{\phi} = (\phi_1, \dots, \phi_{d-1})^T$ be a $(d-1)$ -dimensional

parameter vector and define

$$\boldsymbol{\alpha}(\boldsymbol{\phi}) = \begin{pmatrix} \sqrt{1 - (\phi_1^2 + \dots + \phi_{d-1}^2)} \\ \phi_1 \\ \dots \\ \phi_{d-1} \end{pmatrix}.$$

The true parameter vector $\boldsymbol{\phi}_0$ must satisfy the constraint $\|\boldsymbol{\phi}_0\| < 1$ so that $\boldsymbol{\alpha}(\boldsymbol{\phi})$ is infinitely differentiable in a neighborhood of $\boldsymbol{\phi}_0$. Both of the original constraints $\|\boldsymbol{\alpha}\| = 1$ and $\alpha_1 > 0$ also hold under this reparameterization. If I denote $\boldsymbol{\theta}_\phi = (\boldsymbol{\phi}^T, \boldsymbol{\beta}^T)^T$, it is obvious that $\boldsymbol{\theta}_\phi$ is one-dimensionally lower than $\boldsymbol{\theta}_\alpha = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$. Under this reparameterization, the constrained minimization problem is converted to an unconstrained one. Therefore, I can re-write the corresponding mean function (4.3) as

$$m_{m;i}(\boldsymbol{\theta}_\phi) \triangleq m_m(\mathbf{x}_{m;i}; \boldsymbol{\theta}_\phi) = \boldsymbol{\beta}^T \mathbf{B}(\mathbf{x}_{m;i}^T \boldsymbol{\alpha}(\boldsymbol{\phi})).$$

The gradient vector of the mean function is

$$\dot{m}_{m;i}(\boldsymbol{\theta}_\phi) = \dot{m}_m(\mathbf{x}_{m;i}; \boldsymbol{\theta}_\phi) = \begin{pmatrix} \boldsymbol{\beta}_m^T \dot{\mathbf{B}}(\mathbf{x}_{m;i}^T \boldsymbol{\alpha}(\boldsymbol{\phi})) [-(1 - \|\boldsymbol{\phi}\|^2)^{-1/2} \boldsymbol{\phi} : \mathbf{I}_{d-1}] \mathbf{x}_{m;i} \\ \mathbf{B}^T(\mathbf{x}_{m;i}^T \boldsymbol{\alpha}(\boldsymbol{\phi})) \end{pmatrix}_{\dim(\boldsymbol{\theta}_\alpha) \times 1},$$

where \mathbf{I}_{d-1} is the $(d-1) \times (d-1)$ identity matrix, and $\dot{m}_m(\cdot)$ and $\dot{\mathbf{B}}(\cdot)$ denote first gradient of $m_m(\cdot)$ and $\mathbf{B}(\cdot)$ respectively. The Jacobian matrix of the mapping $\mathcal{F} : \boldsymbol{\theta}_\phi \rightarrow \boldsymbol{\theta}_\alpha$ is

$$\mathbf{J}(\phi) = \boldsymbol{\theta}_\alpha^{(1)}(\boldsymbol{\theta}_\phi) = \begin{pmatrix} -(1 - \|\phi\|^2)^{-1/2} \phi^T & \mathbf{0} \\ \mathbf{I}_{d-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(p+K+1) \times M} \end{pmatrix}_{\dim(\boldsymbol{\theta}_\alpha) \times (\dim(\boldsymbol{\theta}_\alpha) - 1)}.$$

Below, I introduce some notation that is used extensively throughout the proof. I denote $\boldsymbol{\theta}_\phi$ by $\boldsymbol{\theta}$ unless specified explicitly. $\boldsymbol{\theta}_\alpha$ remains its subscript to emphasize its representation of the original parameter space. Similarly, I name $\boldsymbol{\theta}_0 = (\phi_0^T, \beta_0^T)^T$ as the true parameters, while $\boldsymbol{\theta}_{\alpha_0}$ and $\boldsymbol{\theta}_{\phi_0}$ are defined to emphasize the true parameter values in \mathbb{R}^d and \mathbb{R}^{d-1} , respectively. In order to derive the asymptotic properties of estimator, I operate under the following set of assumptions. All of the assumptions are imposed on $\boldsymbol{\theta}$ and the corresponding parameter space Θ .

Assumption 1 *Let Q be a real valued functions on $\Theta \times Y$, where the parameter space Θ is a compact set, and Y is a measurable space. For each $\boldsymbol{\theta}$ in Θ , let $Q(\boldsymbol{\theta}, y)$ be a measurable function of y and the mean function $m(\mathbf{x}; \boldsymbol{\theta})$ is continuous on Θ for each fixed value of \mathbf{x} .*

Assumption 2 *The true parameter vector $\boldsymbol{\theta}_0$ is an interior point of Θ .*

Assumption 3 *The random error vector $\boldsymbol{\epsilon}_i$ satisfies $E\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^T = \boldsymbol{\Sigma}_i$, $\sup_{1 \leq i \leq n} \|\boldsymbol{\Sigma}_i\| \leq \infty$, each with mean zero and moments of some order $(2+r)$, i.e. there exists a constant $r > 0$, such that*

$$\sup_{1 \leq i \leq n, 1 \leq m \leq M} E|\epsilon_{m;i}|^{2+r} < \infty.$$

Assumption 4 Both Σ_i and \mathbf{W}_i are full rank, positive definite matrices with bounded eigenvalues, and

$$e \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \text{tr}(\mathbf{W}_i^{-1} \Sigma_i)$$

exist.

Assumption 5

$$\frac{1}{n} \sum_{i=1}^n \sum_{m=1}^M \left(m_{m;i}(\boldsymbol{\theta}) - m_{m;i}(\boldsymbol{\theta}^*) \right)^2 \quad (4.4)$$

converges uniformly in $\boldsymbol{\theta}, \boldsymbol{\theta}^* \in \Theta$, and if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

$$\mathcal{Q}(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^M \left(m_{m;i}(\boldsymbol{\theta}) - m_{m;i}(\boldsymbol{\theta}_0) \right)^2 \quad (4.5)$$

has a unique minimum.

Assumption 6 The mean function $m(\cdot)$ is twice continuously differentiable in a neighborhood of $\boldsymbol{\theta}_0$, and

$$\Upsilon(\boldsymbol{\theta}_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^M \dot{m}_{m;i}(\boldsymbol{\theta}_0) \dot{m}_{m;i}(\boldsymbol{\theta}_0)^T \quad (4.6)$$

exists and is nonsingular. Furthermore, both

$$\frac{1}{n} \sum_{i=1}^n \sum_{m=1}^M \dot{m}_{m;i}(\boldsymbol{\theta}) \dot{m}_{m;i}(\boldsymbol{\theta})^T \quad (4.7)$$

and

$$\frac{1}{n} \sum_{i=1}^n \sum_{m=1}^M \left[\frac{\partial^2 m_{m;i}(\boldsymbol{\theta})}{\partial \theta_s \partial \theta_t} \right]^2, \quad s, t = 1, \dots, \dim(\boldsymbol{\theta}) \quad (4.8)$$

converge uniformly in $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$.

4.2.2.1 Asymptotics with $\{\lambda_{m;n} \rightarrow 0, \forall m = 1, \dots, M\}$

In this section, each λ_m is denoted by $\lambda_{m;n}$ to indicate the dependence on the sample size n . In order to prove the asymptotic properties of the estimators, I need the following:

Lemma 4.2.1 *Suppose that Assumptions 3 and 4 hold. Then as $n \rightarrow \infty$, we have*

$$\frac{1}{n} \sum_{i=1}^n \boldsymbol{\epsilon}_i^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i \xrightarrow{a.s.} e.$$

Proof Using the fact that the trace of a product are invariant under cyclical permutations of the factors, and that the trace, as a linear operator, commutes with expectation, we have

$$\mathbb{E}(\boldsymbol{\epsilon}_i^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i) = \mathbb{E}\left(\text{tr}(\mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T)\right) = \text{tr}\left(\mathbb{E}(\mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T)\right) = \text{tr}\left(\boldsymbol{\Sigma}_i^{-1} \mathbb{E}(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T)\right) = \text{tr}(\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_i).$$

Define a new set of random variables $X_i = \boldsymbol{\epsilon}_i^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i - \text{tr}(\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_i)$. Obviously, $\{X_i\}$ are independent random variables with mean zero. Assumption states that there exist a constant $r > 0$, such that $\sup_{1 \leq i \leq n, 1 \leq m \leq M} \mathbb{E}|\epsilon_{m;i}|^{2+r} < \infty$, so there must exist $1 < p \leq 2$, such that $\sup_{1 \leq i \leq n, 1 \leq m \leq M} \mathbb{E}|\epsilon_{m;i}|^{2p} < \infty$.

Under assumption 4, There exist constants c_1 and c_2 , such that

$$0 < c_1 \leq \min_{1 \leq i \leq n} \xi_{(1);i} \leq \max_{1 \leq i \leq n} \xi_{(M);i} \leq c_2 < \infty,$$

$$0 < c_1 \leq \min_{1 \leq i \leq n} \zeta_{(1);i} \leq \max_{1 \leq i \leq n} \zeta_{(M);i} \leq c_2 < \infty,$$

where $\xi_{(1);i}$ and $\xi_{(M);i}$ are minimal and maximal eigenvalues of $\boldsymbol{\Sigma}_i$, and $\zeta_{(1);i}$ and $\zeta_{(M);i}$ are minimal and maximal eigenvalues of \mathbf{W}_i , respectively. Therefore, I obtain

$$0 < \boldsymbol{\epsilon}_i^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i \leq \zeta_{(1);i}^{-1} \boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_i \leq c_1^{-1} \boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_i$$

and

$$\begin{aligned}
\text{tr}(\mathbf{W}_i^{-1}\boldsymbol{\Sigma}_i) &\leq \frac{1}{2}\text{tr}(\mathbf{W}_i^{-2}) + \frac{1}{2}\text{tr}(\boldsymbol{\Sigma}_i^2) \\
&= \frac{1}{2}\sum_{m=1}^M \xi_{(m);i}^{-2} + \frac{1}{2}\sum_{m=1}^M \zeta_{(m);i}^2 \\
&\leq \frac{1}{2}\sum_{m=1}^M \xi_{(1);i}^{-2} + \frac{1}{2}\sum_{m=1}^M \zeta_{(M);i}^2 \\
&\leq \frac{1}{2}c_1^{-2}M + \frac{1}{2}c_2^2M.
\end{aligned}$$

Thus, using Cr-inequality, for $p \geq 1$, there exists $c_r = 2^{p-1}$ such that

$$\begin{aligned}
\mathbb{E}|X_i|^p &= \mathbb{E}\left|\boldsymbol{\epsilon}_i^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i - \text{tr}(\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_i)\right|^p \\
&\leq c_r \left(\mathbb{E}|\boldsymbol{\epsilon}_i^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i|^p + \left|\text{tr}(\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_i)\right|^p \right) \\
&\leq c_r \left(\zeta_{(1);i}^{-p} \mathbb{E}|\boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_i|^p + \frac{1}{2} \left|\text{tr}(\mathbf{W}_i^{-2})\right|^p + \frac{1}{2} \left|\text{tr}(\boldsymbol{\Sigma}_i^2)\right|^p \right) \\
&\leq c_r \left(c_1^{-p} \mathbb{E}\left(\sum_{m=1}^M \epsilon_{m;i}^2\right)^p + \frac{1}{2}c_1^{-2}M + \frac{1}{2}c_2^2M \right) \\
&< \infty.
\end{aligned}$$

Thus,

$$\sum_{i=1}^{\infty} \frac{\mathbb{E}|X_i|^p}{i^p} \leq \infty.$$

From Theorem 3.1 of extended version of Chung's Theorem (Chung, 1947; Thanh, 2007)

and also Corollary 8.2 of Kolmogorov's Strong Large Law Number (Klesov, 2014), I get

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{a.s.} 0.$$

Thus, the proof of Lemma 4.2.1 is completed, as $e \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \text{tr}(\mathbf{W}_i^{-1} \boldsymbol{\Sigma}_i)$ exists.

■

Lemma 4.2.2 *Suppose that Assumptions 4 and 6 hold. Then*

$$\begin{aligned}\Psi(\theta_0) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \dot{\mathbf{m}}_i(\theta_0)^T \Sigma_i^{-1} \dot{\mathbf{m}}_i(\theta_0) \\ \Phi(\theta_0) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \dot{\mathbf{m}}_i(\theta_0)^T \mathbf{W}_i^{-1} \Sigma_i^{-1} \mathbf{W}_i^{-1} \dot{\mathbf{m}}_i(\theta_0)\end{aligned}$$

exist and are positive definite matrices.

Proof Using elementary matrix calculus, it is easy to prove that

$$\dot{\mathbf{m}}_i(\theta_0)^T \dot{\mathbf{m}}_i(\theta_0) = \sum_{m=1}^M \dot{m}_{m;i}(\theta_0) \dot{m}_{m;i}(\theta_0)^T,$$

where $\dot{\mathbf{m}}_i(\theta)^T = (\dot{m}_{1;i}(\theta), \dots, \dot{m}_{M;i}(\theta))$ is the transposed gradient matrix of the mean function $\mathbf{m}_i(\theta)$, and $\dot{m}_{m;i}(\theta)$ is defined previously. Because Σ_i^{-1} is a $M \times M$ real symmetric matrix, then Σ_i^{-1} is diagonalizable, and has a set of M linearly independent eigenvectors. Let Δ_i be a square matrix whose columns are those eigenvectors. I can write $\Sigma_i^{-1} = \Delta_i \Lambda_i \Delta_i^T$, where $\Lambda_i = \text{diag}(\xi_{(1);i}^{-1}, \dots, \xi_{(M);i}^{-1})$ is the diagonal matrix such that Λ_{kk} is the eigenvalue associated to column k of Δ_i . Because Δ_i is an orthogonal matrix, there exists a matrix

$$\mathbf{v}_i = \begin{pmatrix} v_{1;i1} & v_{1;i2} & \cdots & v_{1;i(d+p+K+1)} \\ v_{2;i1} & v_{2;i2} & \cdots & v_{2;i(d+p+K+1)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{M;i1} & v_{M;i2} & \cdots & v_{M;i(d+p+K+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{v}_{1;i}^T \\ \mathbf{v}_{2;i}^T \\ \vdots \\ \mathbf{v}_{M;i}^T \end{pmatrix}$$

such that $\dot{\mathbf{m}}_i(\boldsymbol{\theta}_0) = \boldsymbol{\Delta}_i \mathbf{v}_i$. Because $\boldsymbol{\Sigma}_i^{-1}$ is positive definite, $\boldsymbol{\Delta}_i$ is an orthogonal matrix,

$0 < c_1 \leq \min_{1 \leq i \leq n} \xi_{(1);i} \leq \max_{1 \leq i \leq n} \xi_{(M);i} \leq c_2 < \infty$, and

$$\begin{aligned}
\boldsymbol{\Xi}(\boldsymbol{\theta}_0) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^M \dot{m}_{m;i}(\boldsymbol{\theta}_0) \dot{m}_{m;i}(\boldsymbol{\theta}_0)^T \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{v}_i^T \boldsymbol{\Delta}_i^T \boldsymbol{\Delta}_i \mathbf{v}_i \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{v}_i^T \mathbf{v}_i \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^M \mathbf{v}_{m;i} \mathbf{v}_{m;i}^T
\end{aligned}$$

exists and is positive definite. Therefore, I can prove that

$$\begin{aligned}
\boldsymbol{\Psi}(\boldsymbol{\theta}_0) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \boldsymbol{\Sigma}_i^{-1} \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{v}_i^T \boldsymbol{\Delta}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\Delta}_i \mathbf{v}_i \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{v}_i^T \boldsymbol{\Delta}_i^T \boldsymbol{\Delta}_i \boldsymbol{\Lambda}_i \boldsymbol{\Delta}_i^T \boldsymbol{\Delta}_i \mathbf{v}_i \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{v}_i^T \boldsymbol{\Lambda}_i \mathbf{v}_i \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{m=1}^M \xi_{(m);i}^{-1} \mathbf{v}_{m;i} \mathbf{v}_{m;i}^T
\end{aligned}$$

exists and is also positive definite. The existence and positive definite structure of $\boldsymbol{\Phi}(\boldsymbol{\theta}_0)$

can be derived similarly. \blacksquare

Lemma 4.2.3 *Under the same Assumptions listed in Lemma 4.2.2,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \boldsymbol{\Phi}(\boldsymbol{\theta}_0)). \quad (4.9)$$

Proof Let $\gamma = \sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i$. I have

$$\mathbb{E}(\gamma) = \mathbb{E}\left(\sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i\right) = \mathbf{0}$$

and

$$\text{Var}(\gamma) = \text{Var}\left(\sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i\right) = \sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \Sigma_i \mathbf{W}_i^{-1} \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0).$$

For any given unit vector $\mathbf{a} \in \mathbb{R}^{\dim(\boldsymbol{\theta})}$, denote $\mathbf{a}^T \gamma = \sum_{i=1}^n \mathbf{a}^T \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i = \sum_{i=1}^n c_i e_i$, where $c_i^2 = \mathbf{a}^T \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \Sigma_i \mathbf{W}_i^{-1} \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0) \mathbf{a}$, e_i are independent random variables with mean 0 and variance 1. In order to prove (4.9), it suffices to show

$$\frac{1}{n} \sum_{i=1}^n c_i e_i \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \mathbf{a}^T \boldsymbol{\Phi}(\boldsymbol{\theta}_0) \mathbf{a}\right). \quad (4.10)$$

By Lemma 4.2.2, $\boldsymbol{\Phi}(\boldsymbol{\theta}_0)$ exists and is positive definite matrix. I have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n c_i^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{a}^T \left(\sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \Sigma_i \mathbf{W}_i^{-1} \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0) \right) \mathbf{a} > 0.$$

Therefore, the Lindeberg condition holds:

$$\max_{1 \leq i \leq n} \frac{c_i^2}{\sum_{i=1}^n c_i^2} \rightarrow 0.$$

This follows directly from Lemma 3 of Wu (Wu, 1981). Thus, applying *Lindeberg-Feller* central limit theorem, (4.10) is proved, and consequently, (4.9) holds. ■

Theorem 4.2.4 *Under assumptions 1-4, if we further assume that $\lambda_m = o(1), \forall m = 1, \dots, M$, then a sequence of weighted penalized spline least squares estimators $\hat{\boldsymbol{\theta}}_n$ exist*

and is a strongly consistent estimator of $\boldsymbol{\theta}_0$, i.e., as $n \rightarrow \infty$,

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0.$$

Proof After reparameterization, as Assumption 1 holds, the existence of the weighted penalized spline least squares estimators $\hat{\boldsymbol{\theta}}$ follows directly from Lemma 2 in Jennrich's article (Jennrich, 1989).

In the current proof, I retain the notion of λ_m , although it is indeed depending on the sample size. I further denote the weighted penalized spline least squares estimators $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_{n, \lambda_1, \dots, \lambda_M}$ by $\hat{\boldsymbol{\theta}}_{n, \lambda_n^{1:M}}$, thus, it yields minimizing

$$\begin{aligned} Q_{n, \lambda_n^{1:M}}(\boldsymbol{\theta}) &= Q_n(\boldsymbol{\theta}) + \sum_{m=1}^M \lambda_m \boldsymbol{\beta}_m^T \mathbf{D}_m \boldsymbol{\beta}_m \\ &= \frac{1}{n} \sum_{i=1}^n \left(\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\theta}) \right)^T \mathbf{W}_i^{-1} \left(\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\theta}) \right) + \sum_{m=1}^M \lambda_m \boldsymbol{\beta}_m^T \mathbf{D}_m \boldsymbol{\beta}_m \\ &= \frac{1}{n} \sum_{i=1}^n \left(\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\theta}_0) + \mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta}) \right)^T \mathbf{W}_i^{-1} \left(\mathbf{y}_i - \mathbf{m}_i(\boldsymbol{\theta}_0) + \mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta}) \right) \\ &\quad + \sum_{m=1}^M \lambda_m \boldsymbol{\beta}_m^T \mathbf{D}_m \boldsymbol{\beta}_m \\ &= \frac{1}{n} \sum_{i=1}^n \boldsymbol{\epsilon}_i^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i + \frac{2}{n} \sum_{i=1}^n \left(\mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta}) \right)^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left(\mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta}) \right)^T \mathbf{W}_i^{-1} \left(\mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta}) \right) + \sum_{m=1}^M \lambda_m \boldsymbol{\beta}_m^T \mathbf{D}_m \boldsymbol{\beta}_m \\ &= A_1 + A_2 + A_3 + A_4. \end{aligned}$$

$A_1 \xrightarrow{a.s.} e$ follows directly from Lemma 4.2.1. Similar to the proof of Lemma 4.2.2, under the Assumption 2, $\forall \boldsymbol{\theta} \in \Theta$, I can prove that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left(\mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta}) \right)^T \mathbf{W}_i^{-1} \mathbf{W}_i^{-1} \left(\mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta}) \right)$$

exists. Thus, under Assumptions 1 and 2, for almost every ϵ similar to the proof of Theorem 4 of Jennrich (Jennrich, 1989), I have $A_2 \xrightarrow{a.s.} 0$ uniformly $\forall \boldsymbol{\theta} \in \Theta$. Because \mathbf{W}_i is positive definite, I also have

$$A_3 \rightarrow Q(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left(\mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta}) \right)^T \mathbf{W}_i^{-1} \left(\mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta}) \right).$$

Finally, because a set of constant sequence of smoothing parameters $\lambda_m = o(1), \forall m = 1, \dots, M$ and $\boldsymbol{\theta} \in \Theta$ is compact, I have $A_4 \rightarrow 0$. Thus,

$$Q_{n, \lambda_n^{1:M}}(\boldsymbol{\theta}) \xrightarrow{a.s.} Q(\boldsymbol{\theta}) + e \quad (4.11)$$

uniformly $\forall \boldsymbol{\theta} \in \Theta$.

Let $\{\hat{\boldsymbol{\theta}}_{n, \lambda_n^{1:M}}\}$ be a sequence of the weighted penalized spline least squares estimators, and let $\boldsymbol{\theta}'$ be a limit point of such sequence. Let $\{\hat{\boldsymbol{\theta}}_{n_t, \lambda_{n_t}^{1:M}}\}$ be any subsequence of $\{\hat{\boldsymbol{\theta}}_{n, \lambda_n^{1:M}}\}$ which converges to $\boldsymbol{\theta}'$. I want to have $\boldsymbol{\theta}' = \boldsymbol{\theta}_0$.

Under Assumptions 1 and 2 and Equation (4.11), Yu and Ruppert (2002) had shown that $Q_{n_t, \lambda_{n_t}^{1:M}}(\hat{\boldsymbol{\theta}}_{n_t, \lambda_{n_t}^{1:M}}) \xrightarrow{a.s.} Q(\boldsymbol{\theta}') + e$ as $t \rightarrow \infty$. Now, because $\hat{\boldsymbol{\theta}}_{n_t, \lambda_{n_t}^{1:M}}$ is the weighted penalized spline least squares estimate (that is, the minimizer of $Q_{n_t, \lambda_{n_t}^{1:M}}(\cdot)$), then

$$Q_{n_t, \lambda_{n_t}^{1:M}}(\hat{\boldsymbol{\theta}}_{n_t, \lambda_{n_t}^{1:M}}) \leq Q_{n_t, \lambda_{n_t}^{1:M}}(\hat{\boldsymbol{\theta}}_0). \quad (4.12)$$

By letting $t \rightarrow \infty$, I have that the left side of the Equation (4.12) converges to $Q(\boldsymbol{\theta}') + e$, and the right side of the inequality converges to $Q(\boldsymbol{\theta}_0) + e = e$. Thus,

$$Q(\boldsymbol{\theta}') + e \leq e,$$

which yields $Q(\boldsymbol{\theta}') = 0$. Because \mathbf{W}_i is positive definite, by Assumption 2 I can prove that $Q(\boldsymbol{\theta})$ has a unique minimum at $\boldsymbol{\theta}_0$. The limit point $\boldsymbol{\theta}'$ must be $\boldsymbol{\theta}_0$. Because this result holds for almost every ϵ , I have $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}_0$ (i.e. the weighted penalized least squares estimator $\hat{\boldsymbol{\theta}}_n$ is strongly consistent.). ■

Theorem 4.2.5 *Under assumptions 1-6, if we further assume that $\lambda_m = o(n^{-1/2})$, $\forall m = 1, \dots, M$, then as $n \rightarrow \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} MVN\left(\mathbf{0}, \boldsymbol{\Psi}^{-1}(\boldsymbol{\theta}_0) \boldsymbol{\Phi}(\boldsymbol{\theta}_0) \boldsymbol{\Psi}^{-1}(\boldsymbol{\theta}_0)\right),$$

where $\boldsymbol{\Psi}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Phi}(\boldsymbol{\theta}_0)$ are defined in Lemma 4.2.2.

Proof The consistent estimators $\hat{\boldsymbol{\theta}}_n$ minimizes

$$Q_{n, \lambda_n^{1:M}}(\boldsymbol{\theta}) = Q_n(\boldsymbol{\theta}) + \sum_{m=1}^M \lambda_m \boldsymbol{\beta}_m^T \mathbf{D}_m \boldsymbol{\beta}_m.$$

Applying the first Taylor expansion near $\boldsymbol{\theta}_0$, I obtain

$$0 = \left. \frac{\partial Q_{n, \lambda_n^{1:M}}}{\partial \boldsymbol{\theta}} \right|_{\hat{\boldsymbol{\theta}}_n} = \left. \frac{\partial Q_{n, \lambda_n^{1:M}}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}_0} + \left. \frac{\partial^2 Q_{n, \lambda_n^{1:M}}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right|_{\check{\boldsymbol{\theta}}} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

where $\check{\boldsymbol{\theta}}$ is a vector between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$. I denote $\frac{\partial Q_{n, \lambda_n^{1:M}}}{\partial \boldsymbol{\theta}}$ by $\dot{Q}_{n, \lambda_n^{1:M}}$, and $\frac{\partial^2 Q_{n, \lambda_n^{1:M}}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$ by $\ddot{Q}_{n, \lambda_n^{1:M}}$. Then, I have

$$0 = \dot{Q}_{n, \lambda_n^{1:M}}(\hat{\boldsymbol{\theta}}_n) = \dot{Q}_{n, \lambda_n^{1:M}}(\boldsymbol{\theta}_0) + \ddot{Q}_{n, \lambda_n^{1:M}}(\check{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

and, consequently, I have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\ddot{Q}_{n, \lambda_n^{1:M}}(\check{\boldsymbol{\theta}})^{-1} \sqrt{n} \dot{Q}_{n, \lambda_n^{1:M}}(\boldsymbol{\theta}_0). \quad (4.13)$$

Now, I need to show that

$$\sqrt{n}\dot{Q}_{n,\lambda_n^{1:M}}(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \text{MVN}\left(\mathbf{0}, 4\boldsymbol{\Phi}(\boldsymbol{\theta}_0)\right) \quad (4.14)$$

and

$$\ddot{Q}_{n,\lambda_n^{1:M}}(\check{\boldsymbol{\theta}}) \xrightarrow{\mathcal{P}} 2\boldsymbol{\Psi}(\boldsymbol{\theta}_0) \quad (4.15)$$

as $n \rightarrow \infty$.

To prove (4.14), I derive

$$\begin{aligned} \sqrt{n}\dot{Q}_{n,\lambda_n^{1:M}}(\boldsymbol{\theta}_0) &= \sqrt{n}\dot{Q}_n(\boldsymbol{\theta}_0) + (\mathbf{0}_{d-1}^T, 2\sqrt{n}\lambda_1\mathbf{D}_1\boldsymbol{\beta}_{1_0}^T, \dots, 2\sqrt{n}\lambda_M\mathbf{D}_M\boldsymbol{\beta}_{M_0}^T)^T \\ &= -\frac{2}{\sqrt{n}} \sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i + 2\sqrt{n}(\mathbf{0}_{d-1}^T, \lambda_1\mathbf{D}_1\boldsymbol{\beta}_{1_0}^T, \dots, \lambda_M\mathbf{D}_M\boldsymbol{\beta}_{M_0}^T)^T. \end{aligned}$$

The second term on the right side goes to $\mathbf{0}_{\dim(\boldsymbol{\theta})-1}$, because $\lambda_m = o(n^{-1/2})$, $\forall m = 1, \dots, M$.

The summand of the first term on the right side is the weighted average of an independent error sequence. Therefore, by Lemma 4.2.3, I have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i \xrightarrow{\mathcal{D}} \text{MVN}\left(\mathbf{0}, \boldsymbol{\Phi}(\boldsymbol{\theta}_0)\right),$$

where $\boldsymbol{\Phi}(\boldsymbol{\theta}_0)$ is defined in Lemma 4.2.2. Thus, the proof of (4.14) is completed.

To prove (4.15), note that

$$\ddot{Q}_{n,\lambda_n^{1:M}}(\check{\boldsymbol{\theta}}) = \ddot{Q}_n(\check{\boldsymbol{\theta}}) + \text{diag}(\mathbf{0}_{(d-1) \times (d-1)}^T, 2\lambda_1\mathbf{D}_1^T, \dots, 2\lambda_M\mathbf{D}_M^T). \quad (4.16)$$

Again, the second term on the right goes to $\mathbf{0}_{\dim(\boldsymbol{\theta}) \times \dim(\boldsymbol{\theta})}$ because $\lambda_m = o(n^{-1/2})$, $\forall m = 1, \dots, M$. I expand the first term of the right by using expression (2.19) from Amemiya

(Amemiya, 1983):

$$\ddot{Q}_n(\boldsymbol{\theta}) = \frac{2}{n} \sum_{i=1}^n \dot{\mathbf{m}}_i(\boldsymbol{\theta})^T \mathbf{W}_i^{-1} \dot{\mathbf{m}}_i(\boldsymbol{\theta}) - \frac{2}{n} \sum_{i=1}^n \ddot{\mathbf{m}}_i(\boldsymbol{\theta})^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i - \frac{2}{n} \sum_{i=1}^n \ddot{\mathbf{m}}_i(\boldsymbol{\theta})^T \mathbf{W}_i^{-1} (\mathbf{m}_i(\boldsymbol{\theta}_0) - \mathbf{m}_i(\boldsymbol{\theta})),$$

where $\ddot{\mathbf{m}}_i$ denotes $\frac{\partial^2 \mathbf{m}_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$. By Assumption 6, similar to the proof of $A_2 \xrightarrow{a.s.} 0$ uniformly $\forall \boldsymbol{\theta} \in \Theta$, I obtain $\frac{1}{n} \sum_{i=1}^n \ddot{\mathbf{m}}_i(\boldsymbol{\theta})^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i \xrightarrow{a.s.} \mathbf{0}$ uniformly $\forall \boldsymbol{\theta} \in \Theta$. Using the consistency of the weighted penalized spline least squares estimator $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0$, I obtain $\check{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$. This together with Assumption 6 and Lemma 4.2.2 shows that (4.15) holds.

Thus, applying *Slutsky's* lemma to Equation (4.13), Theorem 4.2.5 follows when $\lambda_m = o(n^{-1/2})$, $\forall m = 1, \dots, M$. \blacksquare

Theorem 4.2.6 *Under the same assumptions of Theorem 4.2.5, there exists a sequence of generalized-weighted penalized spline least squares estimators $\hat{\boldsymbol{\theta}}_{\alpha_n} = (\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\beta}}^T)$ with $\|\boldsymbol{\alpha}\| = 1$ which is consistent and asymptotically normally distributed. In other words, as $n \rightarrow \infty$,*

$$\hat{\boldsymbol{\theta}}_{\alpha_n} \xrightarrow{a.s.} \boldsymbol{\theta}_{\alpha_0},$$

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\alpha_n} - \boldsymbol{\theta}_{\alpha_0}) \xrightarrow{\mathcal{D}} MVN\left(\mathbf{0}, \mathbf{J}(\phi_0) \boldsymbol{\Psi}^{-1}(\phi_0) \boldsymbol{\Phi}(\phi_0) \boldsymbol{\Psi}^{-1}(\phi_0) \mathbf{J}(\phi_0)^T\right).$$

Proof Using the results from Theorem 4.2.4 and Theorem 4.2.5 and converting back to the original parameter space via the multivariate delta method, Theorem 4.2.6 follows when $\lambda_m = o(n^{-1/2})$, $\forall m = 1, \dots, M$. \blacksquare

Remark

1. In the above proofs, only fixed-knot asymptotics are considered. Yu and Ruppert (200) argued that this type of asymptotic is more useful than an increasing number of knots for developing a practical statistical methodology. Of course, too few knots will not provide an achievable good fit, whereas too many knots will decrease the efficiency of the model fitting.

Ruppert (Ruppert, 2002) compared minimization of MASE(mean average squared error) over different selections of knot and stated that the default fixed knots is recommended *for effectively all sample sizes* and for all smooth regression functions without too many oscillations. Moreover, asymptotic properties are limited to rates of convergence with an increasing number of knots (Huang, 2003). In my research, fixed knot asymptotics not only give a computationally practical algorithm, but also ensure such convenience to a multivariate normal distribution for inference.

2. In this section, the variance-covariance matrix is assumed unknown which is usually true in practice. An unknown matrix has to be estimated from the data and can affect the estimate of the smoothing parameters (Wang, 1998; Wang et al., 2000). However, when the variance-covariance matrix is known, the existing properties can be easily modified and simplified in theory. Herein, if I assume $\mathbf{W} = \mathbf{\Sigma}$, then the converged covariance in Theorem 4.2.5 will become $\mathbf{\Psi}^{-1}(\boldsymbol{\theta}_0)$ and, consequently, changed to $\mathbf{J}(\boldsymbol{\phi}_0)\mathbf{\Phi}^{-1}(\boldsymbol{\phi}_0)\mathbf{J}(\boldsymbol{\phi}_0)^T$ in Theorem 4.2.6.

3. While the P-spline approach to fitting multivariate single-index models have demonstrated promise in practice and asymptotic properties, I am not optimistic about generalizations to extremely high dimensions. In theory, it is not difficult, but the organization of the computations is difficult. Indeed, the number of the basis functions may easily become larger than the number of observations, and the algorithm may become unstable. Nevertheless, the penalty approach relaxes the importance of the number and location of the knots, and the use of a low-rank smoother solves the computational problems better than other approaches when analyzing large data sets (Durban et al., 2005).

4.2.2.2 Asymptotics with $\{\lambda_{m;n}, \forall m = 1, \dots, M\}$ fixed

The asymptotic normality in 4.2.6 is established when $\lambda_m = o(n^{-1/2}), \forall m = 1, \dots, M$. Therefore, the asymptotic variance does not contain $\{\lambda_m, \forall m = 1, \dots, M\}$. However, if sample size is finite, such asymptotic would overestimate the variance of $\hat{\boldsymbol{\theta}}_{\boldsymbol{\alpha}}$ because some terms are assumed to vanish with infinite samples when approximating the asymptotic variance in Theorem 4.2.5 and Theorem 4.2.6 (Yu and Ruppert, 2002). Thus, for purpose of inference, I provide the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{\boldsymbol{\alpha}}$ when $\{\lambda_{m;n}, \forall m = 1, \dots, M\}$ are fixed.

Recall that in the proof of Theorem 4.2.5,

$$\dot{Q}_{n, \lambda_n^{1:M}} \triangleq \frac{\partial Q_{n, \lambda_n^{1:M}}}{\partial \boldsymbol{\theta}} = -\frac{2}{n} \sum_{i=1}^n \mathbf{m}_i(\boldsymbol{\theta}_0)^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i + 2(\mathbf{0}_{d-1}^T, \lambda_1 \mathbf{D}_1 \boldsymbol{\beta}_{1_0}^T, \dots, \lambda_M \mathbf{D}_M \boldsymbol{\beta}_{M_0}^T)^T,$$

so, the weighted penalized least squares estimator of $\boldsymbol{\theta}$ is the solution of $\dot{Q}_{n, \lambda_n^{1:M}} = 0$.

Similarly, if I define the score function by

$$\mathcal{S}_i(\boldsymbol{\theta}; \lambda_1, \dots, \lambda_M) \triangleq \mathcal{S}_i(\boldsymbol{\theta}; \lambda^{1:M}) = -\mathbf{m}_i(\boldsymbol{\theta})^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i + (\mathbf{0}_{d-1}^T, \lambda_1 \mathbf{D}_1 \boldsymbol{\beta}_1^T, \dots, \lambda_M \mathbf{D}_M \boldsymbol{\beta}_M^T)^T,$$

and solve

$$\sum_{i=1}^n \mathbb{E}\left\{\mathcal{S}_i\left(\boldsymbol{\theta}(\lambda^{1:M}); \lambda^{1:M}\right)\right\} = \mathbf{0}.$$

Then $\hat{\boldsymbol{\theta}}(\lambda^{1:M})$ is unbiased estimator of $\boldsymbol{\theta}(\lambda^{1:M})$. Several articles proposed to estimate the covariance matrix of $\hat{\boldsymbol{\theta}}(\lambda^{1:M})$ by means of sandwich formula. I expand their formula to multivariate cases (Gray, 1994; Yu and Ruppert, 2002). Define

$$\mathbf{L}_i(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\theta}^T} \mathcal{S}_i(\boldsymbol{\theta}) = \frac{1}{2} \ddot{Q}_{n, \lambda_n^{1:M}}(\boldsymbol{\theta}),$$

where $\ddot{Q}_{n,\lambda_n^{1:M}}(\boldsymbol{\theta})$ is defined in (4.16) and

$$\mathbf{R}_i(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{S}_i(\boldsymbol{\theta}) \mathbf{S}_i(\boldsymbol{\theta})^T.$$

Some elementary calculation yields

$$\mathbf{R}_i(\boldsymbol{\theta}) = \mathbf{R}_{1i}(\boldsymbol{\theta}) + \mathbf{R}_{2i}(\boldsymbol{\theta}) + \mathbf{R}_{3i}(\boldsymbol{\theta}),$$

where

$$\mathbf{R}_{1i}(\boldsymbol{\theta}) = \dot{\mathbf{m}}_i(\boldsymbol{\theta})^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^{-1} \mathbf{W}_i^{-1} \dot{\mathbf{m}}_i(\boldsymbol{\theta}),$$

$$\mathbf{R}_{2i}(\boldsymbol{\theta}) = 2\dot{\mathbf{m}}_i(\boldsymbol{\theta})^T \mathbf{W}_i^{-1} \boldsymbol{\epsilon}_i \boldsymbol{\Gamma}_i(\boldsymbol{\theta}),$$

$$\mathbf{R}_{3i}(\boldsymbol{\theta}) = \boldsymbol{\Gamma}_i(\boldsymbol{\theta}) \boldsymbol{\Gamma}_i(\boldsymbol{\theta})^T.$$

Here, $\boldsymbol{\Gamma}_i(\boldsymbol{\theta})$ is a block diagonal matrix, defined by

$$\boldsymbol{\Gamma}_i(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Gamma}_{1i}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{2i}(\boldsymbol{\theta}) \end{pmatrix},$$

where $\boldsymbol{\Gamma}_{1i}(\boldsymbol{\theta}) = \mathbf{0}_{(d-1) \times (d-1)}$ and $\boldsymbol{\Gamma}_{2i}(\boldsymbol{\theta})$ is a $M(p + K + 1) \times M(p + K + 1)$ block matrix.

The (s, t) elements is given by

$$\lambda_s \lambda_t \mathbf{D}_s^T \boldsymbol{\beta}_s \boldsymbol{\beta}_t^T \mathbf{D}_t, \quad , s, t = 1, \dots, M.$$

If I assume both $\frac{1}{n} \mathbf{L}_i(\boldsymbol{\theta})$ and $\frac{1}{n} \mathbf{R}_i(\boldsymbol{\theta})$ converge as $n \rightarrow \infty$, then, similar to the proof of Theorem 4.2.4 and Theorem 4.2.5, I can show that $\hat{\boldsymbol{\theta}}(\lambda^{1:M})$ is a consistent estimator of $\boldsymbol{\theta}(\lambda^{1:M})$ and

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}}(\lambda^{1:M}) - \boldsymbol{\theta}(\lambda^{1:M}) \right) \xrightarrow{\mathcal{D}} \text{MVN} \left(\mathbf{0}, \mathbf{L}_i(\boldsymbol{\theta}(\lambda^{1:M}))^{-1} \mathbf{R}_i(\boldsymbol{\theta}(\lambda^{1:M})) \mathbf{L}_i(\boldsymbol{\theta}(\lambda^{1:M}))^{-T} \right).$$

Therefore, the expression of sandwich estimator $\Phi_{sw}(\hat{\theta}(\lambda^{1:M}))$ for $\text{Cov}(\hat{\theta}(\lambda^{1:M}))$ is given by

$$\Phi_{sw}(\hat{\theta}(\lambda^{1:M})) = \mathbf{L}_i(\hat{\theta}(\lambda^{1:M}))^{-1} \mathbf{R}_i(\hat{\theta}(\lambda^{1:M})) \mathbf{L}_i(\hat{\theta}(\lambda^{1:M}))^{-T}.$$

Similar to the proof of Theorem 4.2.6, the covariance estimator of $\theta_{\alpha}(\lambda^{1:M})$ is followed by

$$\Phi_{sw}(\hat{\theta}_{\alpha}(\lambda^{1:M})) = \mathbf{J}(\hat{\theta}_{\phi}(\lambda^{1:M})) \Phi_{sw}(\hat{\theta}_{\phi}(\lambda^{1:M})) \mathbf{J}(\hat{\theta}_{\phi}(\lambda^{1:M}))^T,$$

where the Jacobian matrix $\mathbf{J}(\cdot)$ is defined previously.

4.3 Simulations

4.3.1 Data Generation

I conduct a simulation study to evaluate the performance of the proposed method in finite sample situations. The simulation is designed as follows. Two correlated outcomes $(\mathbf{y}_1, \mathbf{y}_2)^T$ are generated from the following model for different sample sizes

$$\begin{cases} y_{1,i} = (\alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i})^2 \sin(\alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i}) + \epsilon_{1,i}, \\ y_{2,i} = (\alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i}) \exp(\alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{3i}) + \epsilon_{2,i}, \end{cases}$$

where index component covariates x_{1i} , x_{2i} and x_{3i} are trivariate with independent uniform $[0, 1]$. Two index functions are chosen to be $f_1(v) = v^2 \sin(v)$ and $f_2(v) = v e^v$, respectively. The random error $\epsilon_i = (\epsilon_{1,i}, \epsilon_{2,i})^T \sim \text{MVN}(\mathbf{0}, \Sigma)$, where the correlation coefficient between the two outcomes is ρ , and both σ_1^2 and $\sigma_2^2 = \delta \sigma_1^2$ are used to represent the heteroscedasticity of the two outcomes.

The point estimates for both parameters and variance components are averaged over 200 simulation runs. I use 20 knots to fit cubic spline models. This choice of knots is

enough to resolve the essential structure in the underlying regression functions (Ruppert et al., 2002). The interior knots are computed and selected at equally spaced quantiles of the estimated index values at each iterative step (Wu and Tu, 2012). The smoothing parameters are chosen by ratio of variance components estimated from REML (Krivobokova and Kauermann, 2007; Ruppert et al., 2002).

4.3.2 Simulation Result

Here, results are presented for the case where $\boldsymbol{\alpha} = \frac{1}{\sqrt{14}}(2, -1, 3)^T, \rho = 0.5, \sigma = 2, \delta = 0.9$. Four different sample sizes ($n = 50, 100, 200, 500$) are considered. Again, I compare the estimated values of parameters against the true values. The parameter estimation results, including the mean values of the parameter estimates (Mean), standard error (SE), bias and MSE, are summarized in Table 4.1. Simulation result shows that the estimated coefficient values are close to the true values, and the standard errors estimated based on the multivariate model are consistently smaller. In addition, the empirical standard errors of the estimates of the variance-covariance components closely agree with the true values. Clearly, MSE of each parameter estimates also decrease with increased number of subjects.

Figure 4.1 depicts the average cubic-spline estimates fit to two correlated outcomes with 500 subjects based on 200 simulated datasets, and the corresponding 2.5% and 97.5% quantiles. The behavior of the proposed estimates is superiorly demonstrated. One can see that the bias in the estimated functions are minimal, as the P-spline fits are all close to the true mean functions, although the bias is relatively higher at the tail. Meanwhile, both 2.5% and 97.5% quantiles are close to the true curves, showing very small variation in the estimates.

In summary, the simulation study indicates that the proposed method provides adequate parameter estimates, as well as the ability to capture curvature shape of the true

Table 4.1: Summary of parameter estimates over 200 simulation runs: true parameters $(\alpha_1, \alpha_2, \alpha_3) = \frac{1}{\sqrt{14}}(2, -1, 3) = (0.5345, -0.2673, 0.8018)$, $\rho = 0.50$, $\delta = 0.9$, $\sigma = 2$

Parameter	n	Mean	SE	Bias	MSE
α_1	50	0.5370	0.0008	2.4407e-03	6.6821e-06
	100	0.5349	0.0004	3.7973e-04	2.8654e-07
	200	0.5347	0.0003	1.3222e-04	8.7655e-08
	500	0.5345	0.0002	-5.1607e-05	37761e-08
α_2	50	-0.2659	0.0001	1.3238e-03	3.1092e-06
	100	-0.2656	0.0006	1.6982e-03	3.1892e-06
	200	-0.2678	0.0002	-5.7277e-04	3.7085e-07
	500	-0.2674	0.0002	-1.0535e-04	3.6330e-08
α_3	50	0.8003	0.0004	-1.4751e-03	2.3747e-06
	100	0.8020	0.0003	2.4459e-04	1.4977e-07
	200	0.8015	0.0002	-2.9807e-04	1.2678e-07
	500	0.8018	0.0001	-1.0371e-05	1.7531e-08
ρ	50	0.5223	0.0086	2.2365e-02	5.7429e-04
	100	0.5048	0.0058	4.7662e-03	5.3835e-05
	200	0.5005	0.0037	4.7875e-04	1.4171e-05
	500	0.4946	0.0027	-5.4186e-03	3.6853e-05
δ	50	0.9619	0.0087	6.1932e-02	3.9107e-03
	100	0.9629	0.0059	6.2869e-02	3.9873e-03
	200	0.9289	0.0041	2.8902e-02	8.5243e-04
	500	0.9658	0.0029	6.5827e-02	4.3417e-03
σ	50	1.9571	0.0141	-4.2864e-02	2.0350e-03
	100	1.9715	0.0104	-2.8547e-02	9.2244e-04
	200	2.0119	0.0072	1.1875e-02	1.9230e-04
	500	1.9763	0.0046	-2.3705e-02	5.8338e-04

index functions simultaneously. In addition, variability bands for both estimated curves are notably narrow and close to the nominal level. Finally, MSE of each parameter estimates also apparently decrease with increased sample size. These empirical results agree nicely with the asymptotic properties.

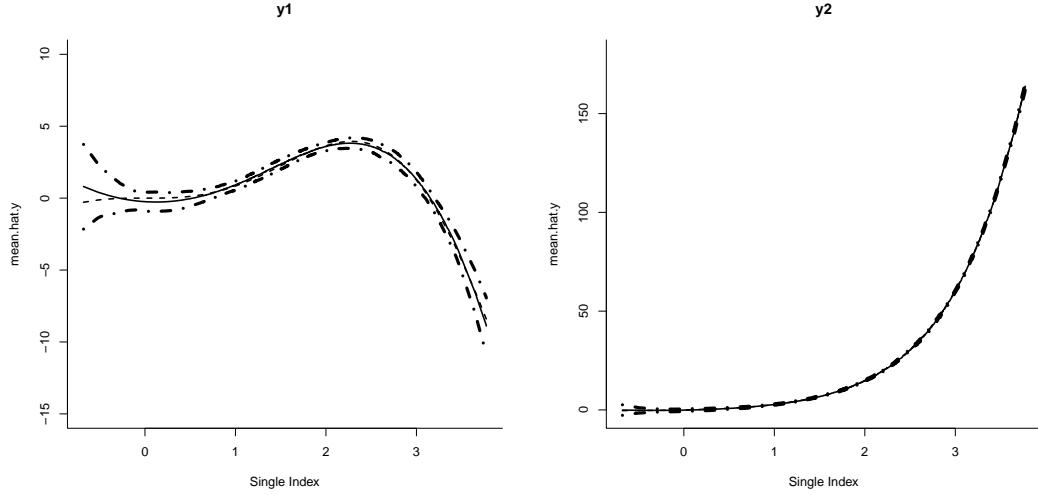


Figure 4.1: Curve estimates and confidence bands for the simulated data with bivariate outcomes. The solid curves are the true mean functions; the dashed curves are the average cubic-spline fit over 200 simulations. The dot-dashed curves are the corresponding 2.5 % and 97.5 % quantiles.

4.4 Discussion

I propose to jointly estimate different single-index models with multiple correlated responses using the weighted penalized least squares. Superior performance of estimators for all parameters are demonstrated in a conducted simulation. Assuming a fixed number of knots, I have shown \sqrt{n} strong consistency and asymptotic normality of the estimators under suitable conditions. From a practical viewpoint, these results will allow us to derive large sample confidence intervals and hypothesis tests and enable us to establish joint inference of all parameters.

Chapter 5

Conclusion

This dissertation discusses several topics related to the single-index regression model in longitudinal and multivariate settings. The overall objective is to develop a set of computationally efficient single-index modeling tools that can be used to derive practically useful indices. I also examine the theoretical properties of the estimator.

The research is motivated by the practical need for data-driven medical indices in epidemiological and clinical investigations. The main methodological thrust of the research is presented in the context of single-index models, which provide a parsimonious representation for high-dimensional factors. The appeal of the discussed approach lies in the intuitive interpretation of monotone index functions. In this research, I extend the traditional single-index model to longitudinal and multivariate data settings, and I adopt the approach for index development. The developed model achieves the ability of dimension reduction by combining multiple factors into a univariate scalar index while retaining nonlinear influences of the predictor variables. With appropriately included covariates, such models also provide an opportunity to estimate index functions while testing other fixed effects. The latter feature helps to distinguish index effects from the effects or confounding effects in clinical investigations.

Fitting a single-index model accounts to optimization with a nonlinear objective function with index components inside the function. Due to the inclusion of an index function structure in the model, index parameters tend to complicate the estimation procedure. To overcome this challenge, I develop a recursive algorithm by using penalized splines with truncated power basis functions for estimation both of index parameters and index function

iteratively. Estimation is performed in a maximum penalized likelihood framework using existing statistical software. Herein, a key step of this approach is to express the model in a mixed effect model, which allowed us to select an optimal smoothing parameter for an approximation of the index function. As a result, the iterative estimation approach is computationally efficient.

Although the classical linear model has a natural extension in a multivariate setting, applying the idea and method of univariate single-index model to multiple longitudinally measured data is not a trivial exercise. Challenges are magnified by the accounting of the two separate sources of correlations, i.e. across-outcome correlation within the same subject, and the temporal correlation within each outcome. Computation associated with such a high-dimensional integration is complicated and challenging. It is for this reason, perhaps, that single-index models for multivariate longitudinal responses has not been widely known and used. Chapter 3 of my dissertation describes a multivariate single-index model that allows for direct estimation of indices and for capturing correlation components across multiple responses. In the proposed model, the two dependency structures are concurrently accommodated using random effect vectors, which, in turn, offset the loss of efficiency in parameter estimation and maintain the estimation consistency. The latent random term not only induces the temporal correlation over repeated collections, but also accounts for dependencies among different outcomes, which adds modeling flexibility and yields unbiased and efficient estimates. To my knowledge, this is the first time that a single-index model for longitudinally multivariate data has been proposed. This approach has been shown to be useful in situations where heterogeneity of the index effects are allowed on each outcome. Similarly, this proposed model may regress index functions on longitudinal data of multiple outcomes. within this framework, parameter estimation can proceed in a traditional mixed

effect model framework; such an approach achieves estimation of multiple parameters in a computationally efficient way.

In chapter 4, I examine the asymptotic properties of the estimators in the proposed model. Reformulating a multivariate single-index model as a mixed effect model gives us an ability to estimate parameters using existing software. In this framework, I explicitly write the weighted penalized likelihood function and show an equivalence of my model to Tian's formulas (Tian et al., 2010). I use this equivalence to derive the asymptotic properties of the estimates including existence, strong consistency, and asymptotic normality. The sandwich estimator of the covariance matrix enables a joint inference of parameters of interest.

In summary, this dissertation bridges the gap between methodological development of single-index models and practical derivation of medical indices. The motivation, importance, and broad application potential of the work is discussed with real data examples. For application, I present examples on the use of these models in analysis of hypertension studies, and to highlight the advantages of analytically derived indices. Simulation studies show that the proposed estimates are both accurate and reliable. The presented data analysis illustrates how an index estimate leads to improved prediction performance. In that regard, I hope this work will enhance the use of single-index regression model in clinical investigations.

BIBLIOGRAPHY

- Amemiya, T. (1983). *Handbook of Econometrics*, Chapter Non-linear Regression Models. Amsterdam, Holland: North-Holland Publishing Company.
- Anestis, A., G. Gérard, and W. M. Ian (2004). Bayesian estimation in single-index models. *Statistica Sinica* 12, 1147–1162.
- Bai, Y., K. Fung, and Y. Zhu (2009). Penalized quadratic inference functions for single-index models with longitudinal data. *Journal of Multivariate Analysis* 100, 152–161.
- Baker, J., L. Olsen, and T. Sorensen (1999). Childhood body-mass index and the risk of coronary heart disease in adulthood. *New England Journal of Medicine* 357, 2329–2337.
- Bellman, R. (1961). *Adaptive Control Processes*. Princeton, NJ: Princeton University Press.
- Brumback, B., D. Ruppert, and M. Wand (1999). Comment on variable selection and function estimation in additive nonparametric regression using a data-based prior by shively, kohn, and wood. *Journal of the American Statistical Association* 94(447), 794–797.
- Carroll, R., J. Fan, I. Gijbels, and M. Wand (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association* 92(483), 477–489.
- Cassano, P., M. Segal, P. Vokonas, and S. Weiss (1990). Body fat distribution, blood pressure, and hypertension: a prospective cohort study of men in the normative aging study. *Ann Epidemiol* 1, 33–48.
- Chung, K. (1947). Note on some strong laws of large numbers. *American Journal of Mathematics* 69, 189–192.

- Cornier, M., J. Despres, N. Davis, D. Grossniklaus, S. Klein, B. Lamarche, F. Lopez-Jimenez, G. Rao, M. St-Onge, A. Towfighi, and P. Poirier (2011). Assessing adiposity : A scientific statement from the American Heart Association. *Circulation* *124*, 1996–2019.
- Crainiceanu, C., D. Ruppert, and M. Wand (2005). Bayesian analysis for penalized spline regression using winbugs. *Journal of Statistical Software* *14*, 1–24.
- Davy, K. and J. Hall (2004). Obesity and hypertension: two epidemics or one? *Circulation* *286*, 803–813.
- Durban, M., J. Harezlak, M. Wand, and R. Carroll (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine* *24*, 1153–1167.
- Dyer, A., K. Liu, M. Walsh, C. Kiefe, and D. J. Jacobs (1999). Ten-year incidence of elevated blood pressure and its predictors: the cardia study. coronary artery risk development in (young) adults. *Journal of Human Hypertension* *13*, 13–21.
- Eilers, P. and B. Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical Science* *11*, 89–121.
- Falkner, B., S. Gidding, G. Ramirez-Garnica, S. Wiltrout, D. West, and E. Rappaport (2006). The relationship of body mass index and blood pressure in primary care pediatric patients. *Pediatr Nephrol* *148*(2), 195–200.
- Gallagher, D., M. Visser, D. Sepulveda, R. Pierson, T. Harris, and S. Heymsfield (1996). How useful is body mass index for comparison of body fatness across age, sex and ethnic groups. *Am J Epidemiol* *143*, 228–239.
- Genton, L., D. Hans, U. Kyle, and C. Pichard (2002). Dual-energy x-ray absorptiometry and body composition: differences between devices and comparison with reference methods. *Nutrition* *18*, 66–70.

- Gray, R. (1994). Spline-based tests in survival analysis. *Biometrics* 50, 640–652.
- Hadaegh, F., M. Bozorgmanesh, M. Safarkhani, D. Khalili, and F. Azizi (2011). Predictability of body mass index for diabetes: Affected by the presence of metabolic syndrome? *BMC Public Health* 11, 1–9.
- Härdle, W., P. Hall, and H. Ichimura (1993). Optimal smoothing in single-index models. *The Annals of Statistics* 21, 157–178.
- Härdle, W. and T. Stoker (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* 84, 986–995.
- Henderson, C. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423–447.
- Hinkle, D., W. Wiersma, and S. Jurs (2003). *Applied Statistics for the Behavioral Sciences*. Boston, MA, U.S.A.: Houghton Mifflin.
- Huang, J. (2003). Asymptotics for polynomial spline regression under weak conditions. *Statistics and Probability Letters* 65, 207–216.
- Huang, Z., W. C. Willett, J. E. Manson, B. Rosner, M. J. Stampfer, F. E. Sprixrt, and G. A. Colditz (1998). Body weight, weight change, and risk for hypertension in women. *Annals of Internal Medicine* 128, 81–88.
- Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Econometrica* 58, 71–120.
- Jennrich, R. (1989). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics* 40, 633–643.
- Klesov, T. (2014). *Limit Theorems for Multi-Indexed Sums of Random Variables*. New York, NY, U.S.A.: Springer.

- Krauss, R. M., M. Winston, B. J. Fletcher, and S. M. Grundy (1998). Impact on cardiovascular disease. *Circulation* 98, 1472–1476.
- Krivobokova, T. and G. Kauermann (2007). A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association* 102, 1328–1337.
- Li, K. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* 86, 316–342.
- MacMahon, S., J. Cutler, E. Brittain, and M. Higgins (1987). Obesity and hypertension: epidemiological and clinical issues. *Eur Heart J* 8(Suppl B), 57–70.
- Marley, J. and M. Wand (2010). Non-standard semiparametric regression via brugs. *Journal of Statistical Software* 37(5).
- Masuo, K., H. Mikami, T. Ogihara, and M. Tuck (2000). Weight gain-induced blood pressure elevation. *Hypertension* 35, 1135–1140.
- Mei, Z., L. Grummer-Strawn, A. Pietrobelli, A. Goulding, M. Goran, and W. Dietz (2002). Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. *Am J Clin Nutr* 75, 978–985.
- NHLBI (1998). Obesity education initiative expert panel on the identification, evaluation, and treatment of overweight and obesity in adults. *National Institutes of Health Publication* 98-4083.
- Pinheiro, J. and D. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. Rensselaer, New York, U.S.A.: Springer.
- Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* 11, 735–757.

- Ruppert, D. and R. Carroll (2000). Spatially adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics* 42, 205–223.
- Ruppert, D., M. Wand, and R. Carroll (2002). *Semiparametric Regression*. Cambridge, U.K.: Cambridge University Press.
- Scherf, J., B. Franklin, C. Lucas, D. Stevenson, and M. Rubenfire (1986). Validity of skinfold thickness measures of formerly obese adults. *The American Journal of Clinical Nutrition* 43, 128–135.
- Sebo, P., S. Beer-Borst, D. Haller, and P. Bovier (2008). Reliability of doctors’ anthropometric measurements to detect obesity. *Preventive Medicine* 47, 389–393.
- Shankar, R. R., G. J. Exkert, C. Saha, W. Tu, Pratt, and J. Howard (2005). The change in blood pressure during pubertal growth. *The Journal of Clinical Endocrinology and Metabolism* 90, 163–167.
- Stamler, J. (1991). Epidemiologic findings on body mass and blood pressure in adults. *Ann Epidemiol* 1, 347–362.
- Steinberger, J., S. Daniels, R. Eckel, L. Hayman, R. Lustig, B. McCrindle, and M. Mietus-Snyder (2009). Progress and challenges in metabolic syndrome in children and adolescents: A scientific statement from the american heart association atherosclerosis, hypertension, and obesity in the young committee of the council on cardiovascular disease in the young; council on cardiovascular nursing; and council on nutrition, physical activity, and metabolism. *Circulation* 119, 628–647.
- Stoker, T. (1986). Consistent estimation of scaled coefficients. *Econometrica* 54, 1461–1481.
- Thanh, L. (2007). On the strong law of large numbers for d-dimensional arrays of random variables. *Electronic Communications in Probability* 12, 434–441.

- Tian, P., L. Yang, and L. Xue (2010). Asymptotic properties of estimators in partially linear single-index model for longitudinal data. *Acta Mathematica Scientia* 30B, 677–687.
- Tu, W., G. Eckert, L. DiMeglio, Z. Yu, J. Jung, and J. Pratt (2011). Intensified effect of adiposity on blood pressure in overweight and obese children. *Hypertension* 58, 818–824.
- Tu, W., G. J. Eckert, C. Saha, and H. J. Pratt (2009). Synchronization of adolescent blood pressure and pubertal somatic growth. *The Journal of Clinical Endocrinology and Metabolism* 94, 5019–5022.
- Wang, J., L. Xue, L. Zhu, and Y. Chong (2010). Partial-linear single-index model with noised variable. *Ann. Statist* 38, 246–272.
- Wang, L. and L. Yang (2009). Spline estimation of single-index models. *Statistics Sinica* 19, 765–783.
- Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of American Statistics Association* 93, 341–348.
- Wang, Y., W. Guo, and M. Brown (2000). Spline smoothing for bivariate data with applications to association between hormones. *Statistica Sinica* 10, 377–397.
- WHO (2000). *Obesity: Preventing and Managing the Global Epidemic - Report on a WHO Consultation*. Geneva, Switzerland: World Health Organization.
- Williams, P. T. (2008). Increases in weight and body size increase the odds for hypertension during 7 years of follow-up. *Obesity* 16, 2541–2548.
- Wu, C. (1981). Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics* 9, 501–513.
- Wu, J. and W. Tu (2012). Development of a pediatric body mass index using longitudinal single-index models. *Statistical Methods in Medical Research*. (In Press).

- Wu, J. and W. Tu (2013). Development of a pediatric body mass index using longitudinal single-index models. *Statistical Methods in Medical Research*. (Epub on January 08, 2013).
- Xia, Y. and W. Hardle (2006). Semi-parametric estimation of partially linear single-index models. *J. Multi. Anal* 97, 1162–1184.
- Xia, Y., W. Härdle, and O. Linton (2012). *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, Chapter Optimal Smoothing for a Computationally and Statistically Efficient Single Index Estimator. Springer-Verlag Berlin Heidelberg: Physica-Verlag HD.
- Xia, Y., H. Tong, W. Li, and L. Zhu (2002). An adaptive estimation of dimension reduction space. *J. R. Statist. Soc. B* 64, 3632–410.
- Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine* 22, 3527–3541.
- Yu, Y. and D. Ruppert (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association* 97(460), 1042–1054.
- Yu, Y. and D. Ruppert (2004). Root-n consistency of penalized spline estimator for partially linear single-index models under general euclidean space. *Statistica Sinica* 14, 449–455.
- Zhang, R. (2007). Tests for nonparametric parts on partially linear single index models. *Science in China Series A: Mathematics* 50, 439–449.
- Zhu, L. and L. Xue (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. Roy. Statist. Soc. ser. B* 68, 549–570.

CURRICULUM VITAE

Jingwei Wu

EDUCATION

- Ph.D. in Biostatistics, Indiana University, Indianapolis, IN, 2015 (minor in Clinical Research)
- M.S. in Biostatistics, University of Michigan, Ann Arbor, MI, 2000
- B.S. Applied Mathematics, Beijing University of Technology, Beijing, P.R.China, 1996

WORKING EXPERIENCE

- Biostatistician, Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN, Aug.2001 - present
- Research Assistant, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, Sep.1998 - Jun.2001
- Network Supervisor, Information Center, China Everbright Group Ltd. , Beijing, P.R.China, Sep.1996 - Jun.1998
- Data Analyst, Institute on Child Statistical Developmental Data, China Children Development Center, Beijing, P.R.China, May.1995 - Sep.1995
- Research Assistant, Department of Applied Mathematics, Beijing University of Technology, Beijing, P.R.China, Sep.1993 - Sep.1994

AWARDS

- Research Assistantship in University of Michigan from 1998 to 2001
- "Outstanding Graduate Award" of Beijing University of Technology in 1996

- Second award of "The National University Students' Computational Mathematics Modeling Competition Outstanding Graduate Award" in 1994
- Undergraduate fellowship in 1994, 1995, 1996

PUBLICATIONS

1. Otte, JL; **Wu, J**; Yu, M; Shaw, C; Carpenter, JS. (2015). "Evaluating the sleep hygiene awareness and practice scale in midlife women with and without breast cancer". *Journal of Nursing Measurement*. (In Press).
2. Bair, MJ; Ang, DC; **Wu, J**; Outcalt, SD; Sargent, C; Kempf, C; Froman, A; Schmid, AA; Damush, TM; YU, Z; Davis, LW; Kroenke, K. (2015). "Evaluation of Stepped Care for Chronic Pain (ESCAPE) in Veterans of Iraq and Afghanistan Conflicts: A Randomized Trial". *JAMA Internal Medicine*. (In Press).
3. Outcalt, SD; Kroenke, K; Krebs, EE; Chumbler, NR; **Wu, J**; Yu, Z; Bair, MJ. (2015). "Chronic Pain and Comorbid Mental Health Conditions: Independent Associations of Posttraumatic Stress Disorder and Depression with Pain, Disability, and Quality of Life". *Journal of Behavioral Medicine*. (In Press).
4. Jalal, S; Hanna, N; Zon, R; Masters, GA; Borghaei, H; Koneru, K; Badve, S; Prasas, N; Somaiah, N; **Wu, J**; Yu, Z. (2014). "Phase I Study of Amrubicin and Cyclophosphamide in Patients With Advanced Solid Organ Malignancies: HOG LUN 07-130". *American Journal of Clinical Oncology*. (In Press).
5. Kroenke, K; Yu, Z; **Wu, J**; Kean, J; Monahan, P. (2014). "Operating characteristics of PROMIS 4-Item depression and anxiety scales in primary care patients with chronic pain". *Pain Medicine*. 15(11): 1892-1901.
6. Kroenke, K; Krebs, EE; **Wu, J**; Yu, Z; Chumbler, NR; Bair, MJ. (2014). "Telecare collaborative management of chronic pain in primary care: a randomized controlled trial". *JAMA*. 312(3): 240-248.

7. Outcalt, SD; Ang, DC; **Wu, J**; Sargent, C; Yu, Z; Bair, MJ. (2014). "Pain experience of Iraq and Afghanistan Veterans with comorbid chronic pain and posttraumatic stress". *Journal of Rehabilitation Research & Development*. 51(4): 559-570.
8. Silbermann, R; Bolzoni, M; Storti, P; Guasco, D; Bonomini, S; Zhou, D; **Wu, J**; Anderson, J; Windle, J; Aversa, F; Roodman, GD; Giuliani, N. (2014). "Bone Marrow Monocyte / Macrophage Derived Activin A Mediates the Osteoclastogenic Effect of IL-3 in Multiple Myeloma". *Leukemia*. 28(4): 951-954.
9. Carpenter, JS; Igega, CM; Otte, JL; Burns, DS; Yu, M; **Wu, J**. (2014). "Somatosensory amplification and menopausal symptoms in breast cancer survivors and midlife women". *Maturitas*. 78(1): 51-55.
10. Shieh, C; **Wu, J**. (2014). "Depressive Symptoms and Obesity/Weight Gain Factors among Black and Hispanic Pregnant Women". *Journal of Community Health Nursing*. 31(1): 8-19.
11. **Wu, J**; Tu, W. (2013). "Development of a Pediatric Body Mass Index Using Longitudinal Single-Index Models". *Statistical Methods in Medical Research*. (In Press).
12. Chumbler, NR; Kroenke, K; Outcalt, SD; Bair, MJ; Kreb, E; **Wu, J**; Yu, Z. (2013). "Association between Sense of Coherence and Health-Related Quality of Life among Primary Care Patients with Chronic Musculoskeletal Pain". *Health and Quality of Life Outcomes*. 11: 216.
13. Brown, LF; Rand, KL; Bigatti, SM; Stewart, JC; Theobald, DE; **Wu, J**; Kroenke, K. (2013). "Longitudinal Relationships Between Fatigue and Depression In Cancer Patients with Depression and/or Pain". *Health Psychology*. 32(12): 1199-1208.
14. Hirsh, AT; Hollingshead, NA; Bair, MJ; Matthias, MS; **Wu, J**; Kroenke, K. (2013). "The influence of patient's sex, race and depression on clinician pain treatment decisions". *European Journal of Pain*. 17(10):1569-1579.

15. Kroenke, K; Outcalt, S; Krebs, E; Bair, MJ; **Wu, J**; Chumbler, N; Yu, Z. (2013).
“Association between anxiety, health-related quality of life and functional impairment
in primary care patients with chronic pain”. *General Hospital Psychiatry*. 35(4): 359-
365.
16. Erler, C; Edwards, NE; Ritchey, S; Pesut, DJ; Sands, L; **Wu, J**. (2013). “Perceived
Patient Safety Culture in a Critical Care Transport Program”. *Air Medical Journal*.
32(4): 208-215
17. Otte, JL; Skaar, TC; **Wu, J**; Yu, M; Ryker, K; Burns, DS; Carpenter, JS. (2013).
“Medication use in breast cancer survivors compared to midlife women”. *Supportive
Care in Cancer*. 21(7): 1827-1833.
18. Kroenke, K; Johns, SA; Theobald, D; **Wu, J**; Tu, W. (2013). “Somatic Symptoms in
Cancer Patients: Trajectory Over 12 Months and Impact on Functional Status and
Disability”. *Supportive Care in Cancer*. 21(3): 765-773.
19. Carpenter, JS; Burns, DS; **Wu, J**; Otte, J; Schneider, B; Ryker, K; Tallman, E; Yu,
M. (2013). “Paced respiration for vasomotor and other menopausal symptoms: a
randomized, controlled trial”. *Journal of General Internal Medicine*. 28(2): 193-200.
20. Kroenke, K; Krebs, EE; **Wu, J**; Bair, MJ; Damush, TM; Chumbler, N; York, T; Weit-
lauf, S; McCalley, S; Evans, E; Barnd, J; Yu, Z. (2013). “Stepped Care to Optimize
Pain Care Effectiveness (SCOPE) Trial: study design and sample characteristics”.
Contemporary Clinical Trials. 34(2): 270-281.
21. Bair, MJ; Poleshuck, EL; **Wu, J**; Krebs, EE; Damush, TM; Tu, W; Kroenke, K.
(2013). “Anxiety but not social stressors predict 12-month depression and pain sever-
ity”. *Clinical Journal of Pain*. 29(2): 95-101.
22. Johns, SA; Kroenke, K; Krebs, EE; Theobald, DE; **Wu, J**; Tu, W. (2013). “Longitu-
dinal comparison of three depression measures in adult cancer patients”. *Journal of
Pain and Symptom Management*. 45(1): 71-82.

23. Carpenter, JS; Burns, DS; **Wu, J**; Yu, M; Ryker, K; Tallman, E; Von Ah, D. (2013). "Strategies used and data obtained during treatment fidelity monitoring". *Nursing Research*. 62(1): 59-65.
24. Yu, M; **Wu, J**; Burns, DS; Carpenter, JS. (2012). "Designing Cost-efficient Randomized Trials by Using Flexible Recruitment Strategies". *BMC Medical Research Methodology*. 12:106.
25. Von Ah, D; Carpenter, JS; Saykin, A; Monahan, PO; **Wu, J**; Yu, M; Rebok, G; Ball, K; Schneider, B; Weaver, M; Tallman, E; Unverzagt, F. (2012). "Advanced cognitive training for breast cancer survivors: A randomized controlled trial". *Breast Cancer Research and Treatment*. 135(3): 799-809.
26. Kroenke, K; Theobald, DE; **Wu, J**; Tu, W; Krebs, EE. (2012). "Comparative responsiveness of pain measures in cancer patients". *Journal of Pain*. 13(8): 764-772.
27. Noureldin, M; Plake, KS; Morrow, DG; Tu, W; **Wu, J**; Murray, MD. (2012). "Effect of Health Literacy on Drug Adherence in Patients with Heart Failure". *Pharmacotherapy*. 32(9): 819-826.
28. Wang, HL; Kroenke, K; **Wu, J**; Tu, W; Theobald, D; Rawl, SM. (2012). "Predictors of Cancer-related Pain Improvement Over Time". *Psychosomatic Medicine*. 74(6): 642-647.
29. Carpenter, JS; **Wu, J**; Burns, DS; Yu, M. (2012). "Perceived control and hot flashes in treatment seeking breast cancer survivors and menopausal women". *Cancer Nursing*. cover article 35(3): 195-202.
30. Kroenke, K; **Wu, J**; Bair, MJ; Damush, TM; Krebs, EE; Tu, W. (2012). "Impact of depression on 12-month outcomes in primary care patients with chronic musculoskeletal pain". *Journal of Musculoskeletal Pain*. 20(1): 8-17.
31. Von Ah, D; Skaar, R; Unverzagt, F; Yu, M; **Wu, J**; Schneider, B; Storniolo, AM; Moser, L; Ryker, K; Milata, J; Carpenter, JS. (2012). "Evaluating the Role of Sero-

- tonin on Neuropsychological Function after Breast Cancer using Acute Tryptophan Depletion”. *Biological Research for Nursing*. 14(1): 5-16.
32. Wang, HL; Kroenke, K; **Wu, J**; Tu, W; Theobald, D; Rawl, SM. (2012). “Cancer-related pain and disability: a longitudinal study”. *Journal of Pain and Symptom Management*. 42(6): 813-821.
 33. Sayar, H; Chan, RJ; Orschell, CM; Chan, EM; Yu, Z; Hood, D; Plett, A; Yang, Z; Hui, CL; Nabinger, SC; Kohlbacher, KJ; West, ES; Walter, A; Sampson, C; **Wu, J**; Cripe, LD. (2012). “Thrice weekly azacitidine does not improve hematological responses in lower-risk myelodysplastic syndromes: a study of the Hoosier Oncology Group”. *Leukemia Research*. 35: 1108-1110.
 34. Gerkensmeyer, JE; Perkins, SM; Day, J; Austin, JK; Scott, EL; **Wu, J**. (2012). “Maternal Depressive Symptoms when Caring for a Child with Mental Health Problems”. *Journal of Child and Family Studies*. 20(5): 685-695.
 35. Kroenke, K; **Wu, J**; Bair, MJ; Krebs, EE; Damush, TM; Tu, W. (2011). “Reciprocal Relationship between Pain and Depression: A 12-Month Longitudinal Analysis in Primary Care”. *The Journal of Pain*. 12(9): 964-973.
 36. Brown, LF; Kroenke, K; Theobald, DE; **Wu, J**. (2011). “Comparison of SF-36 vitality scale and Fatigue Symptom Inventory in assessing cancer-related fatigue”. *Support Care Cancer*. 19: 1255-1259.
 37. Johns, SA; Kroenke, K; Theobald, DE; **Wu, J**; Tu, W. (2011). “Telecare management of pain and depression in patients with cancer: Patient satisfaction and predictors of use”. *Journal of Ambulatory Care Management*. 34(2): 126-139.
 38. Jalal, S; Bedano, P; Einhorn, L; Bhatia, S; Ansari, R; Bechar, N; Koneru, K; Govindan, R; **Wu, J**; Yu, M; Bottema, B; Hanna, N. (2010). “Paclitaxel plus Bevacizumab in Patients with Chemosensitive Relapsed Small Cell Lung Cancer (SCLC): A Safety,

- Feasibility and Efficacy Study (HOG - LUN05-99)". *Journal of Thoracic Oncology*. 5(12); 2008-2011.
39. Casey, EM.; Harb, W; Bradford, D; Bufill, J; Nattam, S; Patel, J; Fisher, W; Latz, JE.; Li, X; **Wu, J**; Hanna, N. "Randomized, Double-Blinded, Multicenter, Phase II Study of Pemetrexed, Carboplatin, and Bevacizumab with Enzastaurin or Placebo in Chemonaïve Patients with Stage IIIB/IV Non-small Cell Lung Cancer: Hoosier Oncology Group (LUN06-116)". *Journal of Thoracic Oncology*. 5(11): 1815-1820.
 40. Kroenke, K; Theobald, D; **Wu, J**; Loza, JK; Carpenter, JS; Tu, W. (2010). "The Association of Depression and Pain with Health-Related Quality of Life, Disability, and Health Care Use in Cancer Patients". *Journal of Pain and Symptom Management*. 40(3): 327-341.
 41. Krebs, EE; Bair, MJ; Damush, TM; Tu, W; **Wu, J**; Kroenke, K. (2010). "Comparative Responsiveness of Pain Outcome Measures Among Primary Care Patients With Musculoskeletal Pain". *Medical Care*. 48(11): 1007-1014.
 42. Kroenke, K; Theobald D; **Wu, J**; Norton, K; Morrison, G; Carpenter, JS; Tu. W. (2010). "Effect of Telecare Management on Pain and Depression in Cancer Patients: A Randomized Trial". *JAMA*. 304(2): 163-171.
 43. Westmoreland, GR; Counsell, SR; Tu, W; **Wu, J**; Litzelman, DK. (2010). "Web-Based Training in Geriatrics for Medical Residents: A Randomized Controlled Trial Using Standardized Patients to Assess Outcomes". *Journal of the American Geriatrics Society*. 58: 1163-1169.
 44. Kroenke, K; Zhong, X; Theobald, D; **Wu, J**; Tu, W; Carpenter, JS. (2010). "Somatic Symptoms in Cancer Patients with Pain and/or Depression: Prevalence, Disability, and Health Care Use." *Archives of Internal Medicine*. 170(18): 1686-1694.

45. Brown, LF; Kroenke, K; Theobald, D; **Wu, J**; Tu, W. (2010). "The association of depression and anxiety with health-related quality of life in cancer patients with depression and/or pain". *Psycho-Oncology*. 19(7): 734-741.
46. Ang, DC; Bair, MJ; Damush, TM; **Wu, J**; Tu, W; Kroenke, K. (2010). "Predictors of Pain Outcomes in Patients with Chronic Musculoskeletal Pain Co-morbid with Depression: Results from a Randomized Controlled Trial". *Pain Medicine*. 11(4): 482-491.
47. Stubbs, D; Krebs, E; Bair, MJ; Damush, TM; **Wu, J**; Sutherland, J; Kroenke, K. (2010). "Sex Differences in Pain and Pain-Related Disability among Primary Care Patients with Chronic Musculoskeletal Pain". *Pain Medicine*. 11(2): 232-239.
48. Baird, CL; Murawski, MM; **Wu, J**. (2010). "Efficacy of Guided Imagery with Relaxation for Osteoarthritis Symptoms and Medication Intake". *Pain Management Nursing*. 11(1): 56-65.
49. Kim, JS; Pressler, SJ; Welch, JL; Damush, TM; Sloan, RS; **Wu, J**; Groh, WJ. (2009). "Physical Function of Patients With Implantable Cardioverter-Defibrillators". *The Journal of Cardiovascular Nursing*. 24(5): 398-409.
50. Jalal, S; Waterhouse, D; Edelman, JM; Nattam, S; Ansari, R; Koneru, K; Clark, R; Richards, A; **Wu, J**; Yu, M; Bottema, B; White, A; Hanna, N. (2009). "Pemetrexed plus Cetuximab in Patients with Recurrent Non-small Cell Lung Cancer (NSCLC): A Phase I/II Study from the Hoosier Oncology Group". *Journal of Thoracic Oncology*. 10(4): 1-5.
51. Carpenter, JS; Yu, M; **Wu, J**; Von Ah, D; Milata, J; Otte, JL; Johns, S; Schneider, B; Storniolo, AM; Salomon, R; Desta, Z; Cao, D; Jin, Y; Philips, S; Skaar, TC. (2009). "Evaluating the role of serotonin in hot flashes after breast cancer using acute tryptophan depletion". *Menopause*. 16(4): 653-659.

52. Krebs, EE; Lorenz, KA; Bair, MJ; Damush, TM; **Wu, J**; Sutherland, JM; Asch, SM; Kroenke, K. (2009). "Development and Initial Validation of the PEG, a Three-item Scale Assessing Pain Intensity and Interference". *Journal of General Internal Medicine*. 24(6): 733-738.
53. Kroenke, K; Bair, MJ; Damush, TM; **Wu, J**; Hoke, S; Sutherland, J; Tu, W. (2009). "Optimized Antidepressant Therapy and Pain Self-management in Primary Care Patients With Depression and Musculoskeletal Pain: A Randomized Controlled Trial". *JAMA*. 301(20): 2099-2110.
54. Murray, MD; Tu, W; **Wu, J**; Morrow, D; Smith, F; Brater, DC. (2009). "Factors Associated With Exacerbation of Heart Failure Include Treatment Adherence and Health Literacy Skills". *Clinical Pharmacology and Therapeutics*. 85: 651-658.
55. Kroenke, K; Theobald, D; Norton, K; Sanders, R; Schlundt, S; McCalley, S; Harvey, P; Iseminger, K; Morrison, G; Carpenter, JS; Stubbs, D; Jacks, R; Carney-Doebbeling, C; **Wu, J**; Tu, W. (2009). "The Indiana Cancer Pain and Depression (INCPAD) trial Design of a telecare management intervention for cancer-related symptoms and baseline characteristics of study participants". *General Hospital Psychiatry*. 31:240-253.
56. Murray, MD; Ritchey, ME; **Wu, J**; Tu, W. (2009). "Effect of a Pharmacist on Adverse Drug Events and Medication Errors in Outpatients With Cardiovascular Disease". *Archives of Internal Medicine*. 169(8):757-763.
57. Pressler, SJ; Gradus-Pizlo, I; Chubinski, SD; Smith, G; Wheeler, S; **Wu, J**; Sloan, R. (2009). "Family Caregiver Outcomes in Heart Failure". *American Journal of Critical Care*. 18(2):149-159.
58. Hansen, RA; Kim, MM; Song, L; Tu, W; **Wu, J**; Murray, MD. (2009). "Comparison of Methods to Assess Medication Adherence and Classify Nonadherence". *Annals of Pharmacotherapy*. 43(3):413-422.

59. Poleshuck, EL; Bair, MJ; Kroenke, K; Damush, TM; Tu, W; **Wu, J**; Krebs, EE; Giles, DE. (2009). "Psychosocial stress and anxiety in musculoskeletal pain patients with and without depression". *General Hospital Psychiatry*. 31:116-122.
60. Duchovic, CA; Gerkenmeyer, JE; **Wu, J**. (2009). "Factors Associated With Parental Distress". *Journal of Child and Adolescent Psychiatric Nursing*. 22(1): 40-48.
61. Westmoreland, GR; Counsell, SR; Sennour, Y; Schubert, CC; Frank, KI; **Wu, J**; Frankel, RM; Litzelman, DK; Bogdewic, SP; Inui, TS. (2009). "Improving Medical Student Attitudes Toward Older Patients Through a "Council of Elders" and Reflective Writing Experience". *Journal of the American Geriatrics Society*. 57: 315-320.
62. Bair, MJ; **Wu, J**; Damush, TM; Sutherland, JM; Kroenke, K. (2008). "Association of Depression and Anxiety Alone and in Combination With Chronic Musculoskeletal Pain in Primary Care Patients". *Psychosomatic Medicine*. 70: 890-897.
63. Damush, TM; **Wu, J**; Bair, MJ; Sutherland, JM; Kroenke, K. (2008). "Self-management practices among primary care patients with musculoskeletal pain and depression". *Journal of Behavioral Medicine*. 31: 301-307.
64. Levinthal, BR; Morrow, DG; Tu, W; **Wu, J**; Murray, MD. (2008). "Cognition and Health Literacy in Patients with Hypertension". *Journal of General Internal Medicine*. 23(8): 1172-1176.
65. Gerkenmeyer, JE; Perkins, SM; Scott, EL; **Wu, J**. (2008). "Depressive Symptoms Among Primary Caregivers of Children With Mental Health Needs: Mediating and Moderating Variables". *Archives of Psychiatric Nursing*. 22(3): 135-146.
66. Kroenke, K; Bair, M; Damush, T; Hoke, SC; Nicholas, G; Kempf, C; Huffman, M; **Wu, J**; Sutherland, J. (2007). "Stepped Care for Affective Disorders and Musculoskeletal Pain (SCAMP) study: Design and practical implications of an intervention for comorbid pain and depression". *General Hospital Psychiatry*. 29: 506-517.

67. Murray, MD; Young, JM; Hoke, SC; Tu, W; Weiner, M; Morrow, DG; Stroupe, KT; **Wu, J**; Clark, DO; Smith, FA; Gradus-Pizlo, I; Weinberger, M; Brater, DC. (2007). "Pharmacist Intervention to Improve Medication Adherence in Heart Failure: A randomized trial". *Annals of Internal Medicine*. 146(10): 714-725.
68. Monahan, PO; Champion, V; Rawl, S; Giesler, RB; Given, B; Given, CW; Burns, D; Bigatti, S; Reuille, KM; Azzouz, F; **Wu, J**; Koch, M. (2006). "What contributes more strongly to predicting QOL during 1-year recovery from treatment for clinically localized prostate cancer: 4-weeks-post-treatment depressive symptoms or type of treatment?". *Quality of Life Research*. 46(5): 669-676.
69. Morrow, D; Clark, D; Tu, W; **Wu, J**; Weiner, M; Steinley, D; Murray, MD. (2006). "Correlates of Health Literacy in Patients with Chronic Heart Failure". *The Gerontologist*. 46(5): 669-676.
70. Bennett, SJ; Litzelman, DK; Wright, A; Perkins, SM; **Wu, J**; Meyer L; Vaughn M. (2006). "The PUMP UP Program for Improving Heart Failure Care: A Feasibility Study". *Nursing Outlook*. 54(1): 39-45.
71. Anand, A; Li, Y; Wang, Y; **Wu, J**; Gao, S; Bukhari, L; Mathews, VP; Kalnin, A; Lowe, MJ. (2005). "Activity and Connectivity of Brain Mood Regulating Circuit In Depression: an fMRI Study". *Biological Psychiatry*. 57: 1079-1088.
72. Anand, A; Li, Y; Wang, Y; **Wu, J**; Gao, S; Bukhari, L; Mathews, VP; Kalnin, A; Lowe, MJ. (2005). "Antidepressant Effect on Connectivity of the Mood Regulating Circuit: an fMRI Study". *Neuropsychopharmacology*. 30: 1334-1344.
73. Subramanian, U; Weiner, M; Gradus-Pizlo, I; **Wu, J**; Tu, W; Murray, MD. (2005). "Patient perception and provider assessment of severity of heart failure as predictors of hospitalization". *Heart and Lung*. 34(2): 89-98.
74. Tu, W; Morris, AB; Li, J; **Wu, J**; Young, JM; Brater, DC; Murray, MD. (2005). "Association between Adherence Measurements of Metoprolol and Healthcare Util-

- lization in Older Heart Failure Patients”. *Clinical Pharmacology and Therapeutics*. 77: 189-201.
75. Fife, KF; Bernstein, DI ; Tu, W; Zimet, GD; Brady, RC; **Wu, J**; Fortenberry, JD; Stone, KM; Rosenthal, RL. (2004). “Predictors of HSV 2 Antibody Positivity among Persons with No History of Genital Herpes”. *Sexually Transmitted Diseases*. 31(11): 676-681. PMID: 15502676
 76. Fortenberry, JD; Zimet, GD; Brady, RC; **Wu, J**; Tu, W; Stone, KM; Rosenthal, SL; Bernstein, DI; JS; Fife, KF. (2004). “Return for Results After Herpes Simplex Virus Type 2 Screening”. *Sexually Transmitted Diseases*. 31(11): 655-658.
 77. Zimet, GD; Rosenthal, SL; Fortenberry, JD; Brady, RC; Tu, W; **Wu, J**; Bernstein, DI; Stanberry, LR; Stone, KM; Leichter, JS; Fife, KF. (2004). “Factors Predicting the Acceptance of Herpes Simplex Virus Type 2 Antibody Testing among Adolescents and Young Adults”. *Sexually Transmitted Diseases*. 31(11): 665-669.
 78. Murray, MD; Young, JM; Morrow, DG; Weiner, M; Tu, W; Hoke, SC; Clark, DO; Stroupe, KT; **Wu, J**; Deer, MM; Bruner-England, TE; Sowinski, KM; Smith, FE; Oldridge, NB; Gradus-Pizlo, I; Murray, LL; Brater, DC; Weinberger, M. (2004). “Methodology of an ongoing random controlled trial of an intervention to improve drug use for elderly patients with chronic heart failure”. *The American Journal of Geriatric Pharmacotherapy*. 2(1): 53-65.
 79. Hope, CJ; **Wu, J**; Tu, W; Murray, MD. (2004). “Association of medication adherence, knowledge, and skills with emergency department visits by adults 50 years or older with congestive heart failure”. *American Journal of Health-System Pharmacy*. 61: 2043-2049.
 80. Hope, CJ; **Wu, J**; Tu, W; Young, J; Murray, MD. (2004). “Barriers to Adherence in Elderly Heart Failure Patients”. *Drug Information Journal*. 38(4): 331-342.

81. Yu, M; **Wu, J**; Maturi, RK. (2003). “Difference of Multifocal Electroretinogram between Progressive Cone Dystrophy and Stargardt’s Disease”. *Chinese Journal of Optometry and Ophthalmology*. 5(3): 129-132 (in English)